SF 298 MASTER COPY   KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
22 May 1998

3. REPORT TYPE AND DATES COVERED
Annual progress, 1 Jun 97 – 1 Jun 98

4. TITLE AND SUBTITLE
Multichannel/Multisensor Signal Processing In Uncertain Environments With Application To Multitarget Tracking

5. FUNDING NUMBERS
ONR G N00014-97-1-0822

6. AUTHOR(S)
Jitendra K. Tugnait

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Auburn University
Dept. of Electrical Engineering
Auburn, AL 36849

8. PERFORMING ORGANIZATION REPORT NUMBER
PR-98-1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Ballston Center Tower One
One North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Navy position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

19980601 082

13. ABSTRACT (*Maximum 200 words*)

This research project is concerned with two distinct aspects of analysis and processing of signals received at multiple sensors from multiple sources when the operating environment is highly uncertain and unstructured. In part I, a general approach based upon an independent component decomposition (ICD) is sought to be investigated involving as few assumptions as possible compared to existing literature. The approach is sought to be developed in conjunction with specific, useful applications such as space and time diversity multiaccess/multiuser digital communications and multitarget tracking using multi-platform multisensor arrays. In part II focus is on maneuvering target tracking using kinematic models. This report describes the progress made on the above two aspects of the project. Details are provided in attached copies of 10 papers – 6 journal articles (accepted/submitted) and 4 conference papers.

14. SUBJECT TERMS
maneuvering target tracking, IMM method, source separation, multiuser interference separation, blind equalization

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

REPORT PR-98-1

ANNUAL PROGRESS REPORT FOR O.N.R. GRANT N00014-97-1-0822

Multichannel/Multisensor Signal Processing In Uncertain Environments With Application To Multitarget Tracking

This research project is concerned with two distinct aspects of analysis and processing of signals received at multiple sensors from multiple sources when the operating environment is highly uncertain and unstructured. In part I, a general approach based upon an independent component decomposition (ICD) is sought to be investigated involving as few assumptions as possible compared to existing literature. The approach is sought to be developed in conjunction with specific, useful applications such as space and time diversity multiaccess/multiuser digital communications and multitarget tracking using multi-platform multisensor arrays. In part II focus is on maneuvering target tracking using kinematic models.

WORK COMPLETED AND IN PROGRESS ("near future") Progress has been made on the following major aspects of the project:

1 **TRACKING MANEUVERING TARGETS USING MULTIPLE KINEMATIC MODELS:** We have investigated a new method (interacting multiple model (IMM) fixed-lag smoothing) for tracking a single maneuvering target in a "clean" environment (no clutter). This work has been reported in [2]. The approach yields much improved performance when compared to filtering at the cost of a slight delay (one or two sampling intervals). We are currently working on the same approach (and its variations) for tracking a single maneuvering target in clutter. We expect to complete a journal paper on it by the end of June 1998. Then we plan to move on to tracking multiple maneuvering targets in clutter.

2 **INDEPENDENT COMPONENT DECOMPOSITION AND ITS APPLICATIONS:** Here we have investigated several approaches for independent source separation, equalization, channel estimation and independent component decomposition. The results have been reported in refs. [1], [3]-[10]. The next step is to focus exclusively on multiple sources and performance analysis.

Journal Articles Submitted/Accepted

[1] J.K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals in unknown additive noise," *IEEE Trans. Signal Processing*, accepted 4/98; not yet scheduled.

[2] B. Chen and J.K. Tugnait, "An interacting multiple model fixed-lag smoothing algorithm for Markovian switching systems," submitted to *IEEE Trans. Aerospace & Electronic Systems.*

[3] J.K. Tugnait and B. Huang, "Second-order statistics-based blind equalization of IIR single-input multiple-output channels with common zeros," submitted to *IEEE Trans. Signal Processing.*

[4] J.K. Tugnait, "Multistep linear predictors-based blind equalization of FIR/IIR single-input multiple-output channels with common zeros," submitted to *IEEE Trans. Signal Processing.*

[5] J.K. Tugnait, "Adaptive blind separation of convolutive mixtures of independent linear signals," submitted to *Signal Processing* (invited paper).

[6] J.K. Tugnait, "Blind adaptive spatio-temporal multiuser signal separation and interference suppression for frequency selective multipath channels," submitted to *IEEE Trans. Communications.*

Conference Presentations & Proceedings Papers

[7] J.K. Tugnait, "Adaptive blind deconvolution of MIMO channels using higher-order statistics," in *Proc. 31st Annual Asilomar Conf. on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 2-5, 1997, pp. 468-472.

[8] B. Huang and J.K. Tugnait, "Blind equalization of I.I.R. single-input multiple-output channels with common zeros using second-order statistics," in *Proc. 1998 IEEE Intern. Conf. Acoustics, Speech, Signal Proc.*, Seattle, WA, May 12-15, 1998, pp. VI-3409-3412.

[9] J.K. Tugnait, "Adaptive blind separation of convolutive mixtures of independent linear signals," in *Proc. 1998 IEEE Intern. Conf. Acoustics, Speech, Signal Proc.*, Seattle, WA, May 12-15, 1998, pp. IV-2097-2100.

[10] J.K. Tugnait, "On multi-step linear predictors for M.I.M.O. F.I.R./I.I.R. channels and related blind equalization," to be presented at *the 1998 IEEE Digital Signal Processing Workshop*, Bryce Canyon, Utah, Aug. 9-12, 1998.

Copies of references [1]-[10] are attached.

[1]

On Blind Separation of Convolutive Mixtures of Independent Linear Signals in Unknown Additive Noise [1] [2]

Jitendra K. Tugnait

Abstract

Blind separation of independent signals (sources) from their linear convolutive mixtures is considered. The various signals are assumed to be linear non-Gaussian but not necessarily i.i.d. First an iterative, normalized higher-order cumulant maximization based approach is exploited using the third- and/or fourth-order normalized cumulants of the "beamformed" data. It provides a decomposition of the given data at each sensor into its independent signal components. In a second approach higher-order cumulant matching is used to consistently estimate the MIMO impulse response via nonlinear optimization. In a third approach higher-order cumulants are augmented with correlations. For blind signal separation the estimated channel is used to decompose the received signal at each sensor into its independent signal components via a Wiener filter. Two illustrative simulation examples are presented.

1 Introduction

Consider a discrete-time FIR MIMO system with N outputs and M inputs given by

$$\mathbf{y}(k) = \sum_{l=0}^{L} \mathbf{F}_l \mathbf{w}(k-l) + \mathbf{n}(k) = [\mathcal{F}(z)]\mathbf{w}(k) + \mathbf{n}(k) = \mathbf{s}(k) + \mathbf{n}(k) \tag{1-1}$$

where $\mathcal{F}(z) = \sum_{i=0}^{L} \mathbf{F}_l z^{-i}$, $\mathbf{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathbf{w}(k)$, $\mathbf{s}(k)$ and $\mathbf{n}(k)$, $w_j(k)$ is the j-th input at sampling time k, $y_i(k)$ is the i-th output, $n_i(k)$ is the additive measurement noise, and $\{\mathbf{F}_l\}$ is the system matrix impulse response (IR). We allow all of the above variables to be complex-valued. We impose the following conditions:

- **(AS1)** $N \geq M$, i.e. there are at least as many outputs as inputs.
- **(AS2)** Rank$\{\mathcal{F}(z)\} = M$ for any $|z| = 1$.
- **(AS3)** The vector sequence $\mathbf{w}(k)$ is assumed to be zero-mean and i.i.d. both temporally and spatially. Also assume that fourth-cumulant or the third-cumulant of $\mathbf{w}(k)$ is nonzero.
- **(AS4)** The noise $\{\mathbf{n}(k)\}$ is a zero-mean, stationary Gaussian sequence (with unknown correlation function) independent of $\{\mathbf{w}(k)\}$. Moreover, it is ergodic.

Let the transfer function of individual subchannels be denoted by $F_{ij}(z)$ (transfer function between the i-th output and j-th input) having the IR $\{f_{ij}(k)\}$.

[1]The author is with the Department of Electrical Engineering, Auburn University, Auburn, AL 36849-5201, USA. E-mail: tugnait@eng.auburn.edu; Tel. (334)844-1846; Fax: (334)844-1809

[2]This work was supported by the National Science Foundation under Grant MIP-9312559 and by the Office of Naval Research under Grant N00014-97-1-0822.

The problem of blind separation of independent linear signals from their convolutive mixtures leads to the above mathematical model. In the convolutive mixture problem, M independent non-Gaussian signals $x_j(k)$ ($j = 1, 2, \cdots, M$) are observed at N sensors as

$$\mathbf{y}(k) = [\mathcal{U}(z)]\mathbf{x}(k) + \mathbf{n}(k) \tag{1-2}$$

where $\mathcal{U}(z)$ represents the convolutive mixture. Assume that

$$\mathbf{x}(k) = [\mathcal{V}(z)]\mathbf{w}(k) \tag{1-3}$$

where $\mathbf{w}(k)$ satisfies (AS3) and $\mathcal{V}(z)$ is diagonal. From (1-2) and (1-3), we obtain (1-1) where $\mathcal{F}(z) = \mathcal{U}(z)\mathcal{V}(z)$ and we have used (if needed) an FIR approximation. Past work on separation of convolutive mixtures may be categorized into several classes: time-domain approaches ([2], [6], [8], [11], [12]), frequency-domain approaches ([3]), adaptive (recursive) approaches ([6], [8], [11]), and non-recursive (batch) approaches ([2], [3], [12]). In this paper we present time-domain non-recursive (batch) approaches. Quite a few of existing approaches are limited either to $M = N = 2$ ([3], [8]) or to $M = N$ ([2],[6], [12]). In this paper we consider a general case of $N \geq M$ with M arbitrary.

Let $\mathcal{F}^{(i)}(z)$ denote the $i-$th column of $\mathcal{F}(z)$. In our formulation of blind convolutive signal separation problem, we are interested in decomposing the observations at various sensors into its independent components, i.e. in estimating $[\mathcal{F}^{(i)}(z)]w_i(k)$ for $i = 1, 2, \cdots, M$ given $\{\mathbf{y}(k)\}$ without having a prior knowledge of $\mathcal{F}(z)$. Our main approach is to first estimate $\mathcal{F}(z)$ (Secs. 2-4) and then estimate $[\mathcal{F}^{(i)}(z)]w_i(k)$ via Wiener filtering (Sec. 5). Others have pursued a different approach as follows. Suppose that there exists a MIMO dynamic system $\mathcal{E}(z)$ with N inputs and M outputs such that the overall $M \times M$ system $\mathcal{T}(z) := \mathcal{E}(z)\mathcal{U}(z)$ decouples the source signals. Following the 2×2 case considered in [3], this implies that we must have ($T_{ij}(z)$ denotes the $ij-$th element of $\mathcal{T}(z)$)

$$\begin{aligned} T_{ij}(z) \quad &= 0 \quad \text{for} \quad i \neq i_j \\ &\neq 0 \quad \text{for} \quad i = i_j \end{aligned} \tag{1-4}$$

where $i = 1, 2, \cdots, M$; $j = 1, 2, \cdots, M$ and $i_j \in \{1, 2, \cdots, M\}$ such that $i_j \neq i_l$ for $j \neq l$. That is, in every column and every row of $\mathcal{T}(z)$ there is exactly one non-zero entry. This approach occurs in the seminal paper [1] and others ([2],[3],[8] and references therein). By discarding all but one of the N entries of the $N-$vector $[\mathcal{F}^{(i)}(z)]w_i(k)$, we can get the solution specified by (1-4). The idea of decomposition of $\mathbf{y}(k)$ into its independent components $[\mathcal{F}^{(i)}(z)]w_i(k)$ to achieve source separation has appeared in [7] using higher-order statistics and in [11] using second-order statistics. In [11] it is required that rank$\{\mathcal{U}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$) whereas our **(AS3)** leads to rank$\{\mathcal{U}(z)\} = M$ only for $|z| = 1$; our examples in Sec. 6 do not satisfy the assumptions of [11]. On the other hand, [11] does not require the signals $\{\mathbf{x}(k)\}$ to be non-Gaussian or linear whereas our formulation relies crucially on $\{\mathbf{x}(k)\}$ being linear non-Gaussian.

The assumption of linear non-Gaussian sources allows one to treat the problem as a (blind) linear system identification problem using higher-order statistics. Therefore, existing results on blind system

identification ([4], [5], [12] etc.) become quite relevant. In [5] (also [6]) a source-iterative, inverse-filter criteria-based approach has been developed. It was shown in [5] that the system matrix IR sequence $\{\mathbf{F}_l\}$ can be found up to a post-multiplication monomial matrix. The approach of [5] does not require knowledge of the model order (L in (1-1)). However, it yields biased IR estimates in noise. One of the purposes of this paper is to investigate alternative approaches to remedy this drawback so that consistent channel estimates may be used for source separation given dynamic mixtures. We propose to use a cumulant matching approach [4]. Since higher-order cumulants of Gaussian processes vanish, cumulant matching has the potential of yielding unbiased estimates. Using output cumulants, closed-form solutions have been given in [4] (and references therein) under several restrictive conditions: given a finite impulse response $\{\mathbf{F}_l\}_{l=0}^{L}$, model order L is known, and $\mathbf{F}_0$ and $\mathbf{F}_L$ are both of full column rank. Quadratic cumulant matching has also been performed in [4] under the same restrictive conditions. Since cumulant matching results in a nonlinear optimization problem, selection of good initial guesses is crucial. In [4] this is accomplished by using the closed-form solution. In [12] it has been shown that if two models have identical output cumulants, then their transfer functions (hence impulse responses) are equivalent up to a monomial matrix. However, [12] offers no algorithms for model identification. In this note we utilize the results of the approach of [5] as an initial guess for cumulant matching and related approaches (see further remarks in Sec. 3.). Once the system IR has been estimated, we design an MMSE (minimum mean-square error) filter for signal separation in Sec. 5 using the estimated IR.

2 An Iterative Solution Based on Inverse-Filter Criteria [5]

Here we briefly review [5] whose analysis holds only for $\mathbf{n}(k) \equiv 0$. Let $\mathrm{CUM}_4(w)$ denote the fourth-order cumulant of a complex-valued random variable w, defined as

$$\mathrm{CUM}_4(w) := \mathrm{cum}_4\{w, w^*, w, w^*\} = E\{|w|^4\} - 2[E\{|w|^2\}]^2 - |E\{w^2\}|^2. \tag{2-1}$$

We will use the notation $\gamma_{4wi} = \mathrm{CUM}_4(w_i(k))$ and $\sigma^2_{wi} = E\{|w_i(k)|^2\}$. Consider an $1 \times N$ row-vector polynomial equalizer $C^T(z)$, with its j-th entry denoted by $C_j(z)$, operating on the data vector $\mathbf{y}(k)$. Let the equalizer output be denoted by $e(k) = \sum_{i=1}^{N} C_i(z) y_i(k)$. Following [5] consider maximization of the cost (an inverse-filter criterion)

$$J_{42} := |\mathrm{CUM}_4(e(k))| \times [E\{|e(k)|^2\}]^{-2} \tag{2-2}$$

for designing a linear equalizer to recover one of the inputs. It is shown [5] that when (2-2) is maximized w.r.t. $C(z)$, then $e(k)$ is given by

$$e(k) = d w_{j_0}(k - k_0) \tag{2-3}$$

where d is some complex constant, k_0 is some integer, j_0 indexes some input out of the given M inputs, i.e., the equalizer output is a possibly scaled and shifted version of one of the system inputs. It has been established in [5] that under **(AS1)**-**(AS3)** and no noise, such a solution exists and if doubly-infinite equalizers are used, then all locally stable stationary points of the given cost w.r.t. the equalizer coefficients are also characterized by solutions such as (2-3).

An iterative solution where we iterate on input sequences one-by-one is summarized in Table 1. In practice, all the expectations in (T-1) are replaced with their sample averages over appropriate data records. It has been shown in [4] that

$$\widetilde{y}_{i,j_0}(k) = \sum_{l} f_{ij_0}(l) w_{j_0}(k-l) \tag{2-4}$$

representing the contribution of $\{w_{j_0}(k)\}$ to the i–th sensor: **blind signal separation.**

Remark 1. We may replace the cost (2-2) with ([5]) $J_{32} := |\text{CUM}_3(e(k))|[E\{|e(k)|^2\}]^{-1.5}$ where $\text{CUM}_3(w) := \text{cum}_4\{w, w^*, w\} = E\{|w|^2 w\}$. The preceding discussion pertaining to (2-2) holds in this case with obvious modifications provided we replace the phrase "nonzero fourth cumulants" in (**AS3**) with the phrase "nonzero third cumulants." □

Remark 2. It has been shown in [5] that under the conditions (**AS1**)-(**AS3**) and no noise, the proposed iterative approach yields a transfer function $\mathcal{A}(z)$ which is related to $\mathcal{F}(z)$ via

$$\mathcal{A}(z) = \mathcal{F}(z)\mathbf{DAP} \tag{2-5}$$

where $\mathbf{D}$ is an $M \times M$ "time-shift" diagonal matrix, $\mathbf{A}$ is an $M \times M$ diagonal scaling matrix, and $\mathbf{P}$ is an $M \times M$ permutation matrix □

3 Cumulant Matching

Define

$$C_{ijkl}(\tau_1, \tau_2, \tau_3) := \text{cum}_4\{y_i(t), y_j^*(t+\tau_1), y_k(t+\tau_2), y_l^*(t+\tau_3)\}. \tag{3-1}$$

Let $C_{ijkl}(\tau_1, \tau_2, \tau_3|\theta)$ denote the relevant variable parametrized by θ where θ denotes the vector of all unknown parameters composed of the elements of $\mathbf{F}_l$ for $l = 0, 1, \cdots, \overline{L} \geq L$. Furthermore, let $\widehat{C}_{ijkl}(\tau_1, \tau_2, \tau_3)$ denote a consistent data-based estimate of $C_{ijkl}(\tau_1, \tau_2, \tau_3)$ obtained by appropriate sample averaging. It is easily seen that for (1-1),

$$C_{ijkl}(\tau_1, \tau_2, \tau_3|\theta) = \sum_{m=1}^{M} \sum_{t=0}^{\overline{L}} \gamma_{4wm} f_{im}(t) f_{jm}^*(t+\tau_1) f_{km}(t+\tau_2) f_{lm}^*(t+\tau_3). \tag{3-2}$$

The cost function for parameter estimation via cumulant matching is given by

$$\mathcal{R}_4 := \sum_{i,j,k,l=1}^{N} \sum_{\tau_1=0}^{\overline{L}} \sum_{\tau_2=0}^{\overline{L}} \sum_{\tau_3=0}^{\tau_1} \left| \widehat{C}_{ijkl}(\tau_1, \tau_2, \tau_3) - C_{ijkl}(\tau_1, \tau_2, \tau_3|\theta) \right|^2. \tag{3-3}$$

During minimization of (3-3), γ_{4wm} (see (3-2)) is kept fixed at its value obtained from Sec. 2 using (2-3). This indirectly fixes the scale ambiguity ($\mathbf{A}$ in (2-5)). The initial values of θ are provided by the solution of Sec. 2. The choice of lags in (3-3) reflects the non-redundant region of support for cumulants of complex FIR processes [13]. Minimization of (3-3) can be performed using gradient-based methods (as in [9]) and/or using software packages. For the results presented in Sec. 6 we used NL2SOL [14] with the option of numerical gradients so that explicit equations for gradients were not used.

Remark 3. With $\widehat{C}_{ijkl}(\tau_1,\tau_2,\tau_3)$ replaced with its true value $C_{ijkl}(\tau_1,\tau_2,\tau_3|\theta_0)$ in (3-3), it follows from [12] that under the conditions (**AS1**)-(**AS4**), minimization of (3-3) (under $\overline{L} \geq L$) will yield a transfer function $\mathcal{A}(z)$ which is related to the true transfer function via (2-5). If $\widehat{C}_{ijkl}(\tau_1,\tau_2,\tau_3)$ is a strongly consistent estimator of $C_{ijkl}(\tau_1,\tau_2,\tau_3|\theta_0)$, then (global) minimizer of (3-3) will yield with probability one a transfer function satisfying (2-5) [4]. The problem is how to ensure global minimization of (3-3). Herein lies the value of the iterative approach of Sec. 2. Recall that the approach of Sec. 2 yields a consistent IR estimator only under vanishing measurement noise.

4 Correlation and Cumulant Matching

Let $\overline{\theta}$ denote θ of Sec. 3 augmented with $E\{|w_i(k)|^2\}$ $(=\sigma_{wi}^2)$, $i = 1,2,\cdots,M$. Define $C_{ij}(\tau) :=$ $E\{y_i(t+\tau)y_j^*(t)\}$ and let $C_{ij}(\tau|\overline{\theta})$ denote $C_{ij}(\tau)$ parametrized by $\overline{\theta}$. Then $C_{ij}(\tau|\overline{\theta}) = \sum_{m=1}^{M}\sum_{t=0}^{\overline{L}}\sigma_{wm}^2 f_{im}(t+\tau)f_{jm}^*(t)$. Let $\widehat{C}_{ij}(\tau)$ denote a consistent data-based estimate of $C_{ij}(\tau)$. The cost (3-3) may be augmented with correlation matching to devise the cost

$$\mathcal{R}_{42} = \mathcal{R}_4 + \lambda \sum_{i,j=1}^{N}\sum_{\tau=1}^{\overline{L}} |\widehat{C}_{ij}(\tau) - C_{ij}(\tau|\overline{\theta})|^2 \tag{4-1}$$

where the nonnegative scalar λ in (4-1) is chosen to provide relative weighting between correlation and cumulant matching (as in [9] and [13] for scalar systems). Following [9] we choose

$$\lambda = \lambda_0 \left[\sum_{i,j,k,l=1}^{N}\sum_{\tau_1=0}^{\overline{L}}\sum_{\tau_2=0}^{\overline{L}}\sum_{\tau_3=0}^{\tau_1}\left|\widehat{C}_{ijkl}(\tau_1,\tau_2,\tau_3)\right|^2\right]\left[\sum_{i,j=1}^{N}\sum_{\tau=0}^{\overline{L}}\left|\widehat{C}_{ij}(\tau)\right|^2\right]^{-1} \tag{4-2}$$

where $\lambda_0 \geq 0$. By (4-2) λ is invariant to any scaling of the data. For simulation results presented later, we picked $\lambda_0 = 1$. The initial values of parts of $\overline{\theta}$ that are common to θ are selected as in Sec. 3. The initial values of $\sigma_{wj_0}^2$ are obtained using (2-3) in a manner similar to that in Step (i) of Table 1. The cost (4-1) is useful when noise is white Gaussian (notice the exclusion of $\tau = 0$ in (4-1)). It allows us to exploit the signal correlations at nonzero lags. It can be easily modified to incorporate a different prior knowledge such as known noise correlation etc.

5 Blind Convolutive Signal Separation

As noted earlier, our objective is to estimate $[\mathcal{F}^{(i)}](z)w_i(k)$ for $i = 1,2,\cdots,M$ given $\{\mathrm{y}(k)\}$. The solution of Sec. 2 provides a solution in the form of (2-4) but it is not necessarily an MMSE solution. We will now discuss other possible solutions, particularly when cumulant (or related) matching is used to estimate the (over-all) system transfer function. Let $\mathrm{F}_l^{(i)}$ denote the i−th column of F_l. Let $\widehat{\mathcal{F}}^{(i)}(z)$ and $\widehat{\mathrm{F}}_l^{(i)}$ denote the estimates of $\mathcal{F}^{(i)}(z)$ and $\mathrm{F}_l^{(i)}$, respectively. We wish to design a linear MMSE filter $\{\mathrm{G}_i\}_{i=0}^{L_e}$ of length L_e+1 to estimate $\tilde{\mathrm{y}}^{(j)}(k-d)$ given $\mathrm{y}(l)$ for $l = k, k-1,\cdots,k-L_e$ where $d \geq 0$,

$$\tilde{\mathrm{y}}^{(j)}(k) := [\mathcal{F}^{(j)}(z)]w_j(k) = \sum_{l=0}^{\overline{L}}\mathrm{F}_l^{(j)}w_j(k-l), \tag{5-1}$$

$$\hat{\tilde{\mathrm{y}}}^{(j)}(k-d) := \sum_{i=0}^{L_e}\mathrm{G}_i\mathrm{y}(k-i). \tag{5-2}$$

Both L_e and the delay d are "pre-determined." Using the orthogonality principle [10], the normal equations for the MMSE estimator simplify to

$$\sum_{i=0}^{L_e} \mathbf{G}_i \mathbf{R}_{yy}(m-i) \;=\; \sigma_{wj}^2 \sum_{k=0}^{\overline{L}} \mathbf{F}_k^{(j)} \mathbf{F}_{k+d-m}^{(j)\mathcal{H}} \;=\; \sigma_{wj}^2 \mathbf{H}_{d-m}, \quad m = 0, 1, \cdots, L_e \tag{5-3}$$

where $\mathbf{R}_{yy}(m) := E\{\mathbf{y}(t+m)\mathbf{y}^{\mathcal{H}}(t)\}$ ($\mathcal{H}$ denotes the Hermitian operation) and

$$\mathbf{H}_{d-m} := \sum_{k=0}^{\overline{L}} \mathbf{F}_k^{(j)} \mathbf{F}_{k+d-m}^{(j)\mathcal{H}} = \sum_{k=0}^{\overline{L}} \mathbf{F}_{k+m-d}^{(j)} \mathbf{F}_k^{(j)\mathcal{H}} = \sigma_{wj}^{-2} E\left\{\tilde{\mathbf{y}}^{(j)}(k+m-d)\tilde{\mathbf{y}}^{(j)\mathcal{H}}(k)\right\}. \tag{5-4}$$

Note that a shift in the sequence $\{\mathbf{F}_k^{(j)}\}$ leaves $\mathbf{H}_{d-m}$ unaffected. In order to obtain a data-based solution, we simply replace all the unknowns by their estimates. Since there is an inherent scale ambiguity in estimating the composite channel impulse response (cf. (2-5)), we design the equalizer only up to a scale factor by omitting σ_{wj}^2 from (5-3). Denoting the so modified equalizer gains as $\widehat{\mathbf{G}}_i$ (instead of $\mathbf{G}_i$), we have the solution

$$\left[\; \widehat{\mathbf{G}}_0 \quad \widehat{\mathbf{G}}_1 \quad \cdots \quad \widehat{\mathbf{G}}_{L_e} \;\right] = \left[\; \widehat{\mathbf{H}}_d \quad \widehat{\mathbf{H}}_{d-1} \quad \cdots \quad \widehat{\mathbf{H}}_{d-L_e} \;\right] \widehat{\mathcal{R}}_{yy}^{-1} \tag{5-5}$$

where $\widehat{\mathbf{H}}_{d-m} := \sum_{k=0}^{\overline{L}} \widehat{\mathbf{F}}_k^{(j)} \widehat{\mathbf{F}}_{k+d-m}^{(j)\mathcal{H}}$, $\widehat{\mathbf{R}}_{yy}(m) := T^{-1}\sum_{t=1}^{T} \mathbf{y}(t+m)\mathbf{y}^{\mathcal{H}}(t)\}$, T =record length and $\left[\widehat{\mathcal{R}}_{yy}\right]_{ij} = \widehat{\mathbf{R}}_{yy}(j-i) = ij-$th block of $\widehat{\mathcal{R}}_{yy}$. We assume that noise is such that the inverse in (5-5) exists, else a pseudo-inverse is warranted. The estimates $\widehat{\mathbf{F}}_l^{(i)}$ above may be obtained by any of the previous approaches resulting in several possible choices. Under **(AS3)**-**(AS4)**, $\widehat{\mathbf{R}}_{yy}(m)$ is a consistent estimator of $\mathbf{R}_{yy}(m)$. Therefore, if $\widehat{\mathbf{F}}_l^{(i)}$ is a consistent estimator (to within the ambiguities specified in (2-5)), then asymptotically we have the desired MMSE linear equalizer within a scale factor. This holds true for the approaches of Secs. 3 and 4, but not for that of Sec. 2.

6 Simulation Examples

We now present two simulation examples. In both the examples $\mathbf{F}_0$ is of rank $1 < M = 2$. Calculation of $\widehat{\mathcal{R}}_{yy}^{-1}$ (cf. (5-5)) was performed via singular value decomposition where all singular values $<$ [0.001$\times$(largest singular value)] were neglected. This results in a pseudo-inverse. The various performance measures used (and their computational details) are shown in Table 2. Nonlinear optimization was done using NL2SOL with numerical gradients [14].

Example 1. Consider a 2-input 3-output MA(2) system model resulting in N=3 and M=2 in (1-1). Its 3×2 transfer function $\mathcal{F}(z)$ was chosen as

$$\begin{bmatrix} 0.9078 + 0.9078z^{-2} & 0.7471 + 1.1206z^{-1} + 0.7471z^{-2} \\ 0.7263z^{-1} - 0.9078z^{-2} & 0.5603z^{-1} - 0.5603z^{-2} \\ 0. & 0. \end{bmatrix}. \tag{6-1}$$

The last row of (6-1) is identically zero signifying that the third 'sensor' is not receiving any information signal, just noise. The inputs $\{w_j(k)\}$ ($j = 1, 2$) are mutually independent, zero-mean and i.i.d. such that $w_1(k)$ is one-sided exponential with variance 0.64, and $w_2(k)$ is binary taking values ± 1.0 with

probability 0.5 each. The noise at the three sensors is mutually independent, zero-mean white Gaussian such that the noise power at the first sensor is nine times the noise power at the other two sensors, the latter being equal. Fig. 1 shows the "subchannel" amplitude spectra for the non-null subchannels where ij−th subchannel refers to $F_{ij}(z)$.

The source-iterative approach of Table 1 was applied to inverse filter the data, to estimate the system IR, and to carry out signal separation. The length of the inverse filters was 15 samples per sensor/output. The average signal-to-noise ratio (SNR $= N^{-1}\sum_{i=1}^{N}\left[E\{|s_i(k)|^2\}/E\{|n_i(k)|^2\}\right]$, $s_i(k) = i$−th component of $\mathbf{s}(k)$ in (1-1)) was taken to be 30 dB, 20 dB, 10 dB and 5 dB, respectively, in two sets of 50 Monte Carlo runs with varying record lengths of 1500 and 9000 samples per run, respectively. The results of this approach were used to initialize minimization of (3-3) with $\overline{L} = 3$ and also of (4-1) with $\overline{L} = 3$ and $\lambda_0 = 1$. The channel estimation errors (NMSE) are shown in Fig. 2. The average SINR (=(SINR$_1$ + SINR$_2$)/2) values are shown in Figs. 3 and 4. To design the MMSE equalizer (5-5) we took $L_e = 14$ (as for inverse filters in Table 1) and $d = 7$ in all cases. The approach labeled "inverse filter criterion" in Figs. 3 and 4 uses (T-2) for source separation; other approaches use the MMSE filter of Sec. 5 based upon the estimated channel. It is seen that the inverse filter criteria based approach of Sec. 2 coupled with the MMSE filter with delay $d = 7$ performs quite well for signal separation at higher SNR's. At lower SNR's, cumulant matching does better. The benefits of introducing a delay in signal separation are clear from Figs. 3 and 4. The upper bounds shown in Figs. 3 and 4 were obtained by using the true values of $\mathbf{F}_l^{(j)}$ in (5-4) and estimated $\mathcal{R}_{yy}$ for upper bound (est. cor.) and true $\mathcal{R}_{yy}$ for upper bound (true cor.).

Example 2. Consider a 2-input 3-output MA(6) system model resulting in N=3 and M=2 in (2-1). Its 3×2 transfer function $\mathcal{F}(z)$ was chosen as

$$\begin{bmatrix} 0.7426 + 0.7426z^{-2} & 0.5678 + 0.3407z^{-1} \\ 0.4456z^{-1} + 0.7426z^{-2} & -0.2385z^{-1} - 0.5678z^{-2} + 0.8176z^{-3} + 0.4088z^{-4} + 0.2385z^{-6} \\ 0.8911z^{-2} + 0.5941z^{-3} & 0.6814z^{-1} + 0.9085z^{-2} \end{bmatrix}. \tag{6-2}$$

This example has been taken from [5]. The inputs $\{w_j(k)\}$ $(j = 1, 2)$ are mutually independent, zero-mean and i.i.d. such that $w_1(k)$ takes values ± 0.8 with probability 0.5 each, and $w_2(k)$ takes values ± 1.0 with probability 0.5 each. The noise at the three sensors is mutually independent, zero-mean white Gaussian such that the noise power at all the sensors is the same. Fig. 5 shows the "subchannel" amplitude spectra. The simulation results are shown in Figs. 6–8 using the same procedure and parameters as that for Example 1 except that now we take $\overline{L} = 7$. It is seen that the inverse filter criteria based approach of Sec. 2 coupled with the MMSE filter performs quite well for signal separation at higher SNR's.

7 Conclusions

The problem of blind separation of independent linear non-Gaussian signals from their linear convolutive mixtures observed in additive Gaussian noise of unknown correlation function was considered. Emphasis

was on a two-step procedure where first we estimate the system IR (using one of three approaches) and then design an MMSE filter with a controlled delay for signal separation based upon the estimated IR. Two simulation examples were presented where it was found that the introduction of the delay in MMSE filter design significantly improved the separation performance at the expense of increased computational complexity.

8 References

[1] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromorphic architecture," *Signal Processing*, vol. 24, pp. 1-10, 1991.

[2] J.L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs," *IEEE Trans. Signal Processing*, vol. SP-40, pp. 3074-3078, Dec. 1992.

[3] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Trans. Signal Processing*, vol. SP-42, pp. 2158-2168, Aug. 1994.

[4] A. Swami, G.B. Giannakis and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Proc.*, vol. SP-42, pp. 898-914, April 1994.

[5] J.K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Proc.*, vol. SP-45, pp. 658-672, March 1997.

[6] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. Intern. Conf. Commun.*, pp. 1160-1164, Geneva, Switzerland, June 1993.

[7] J.K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals," in *Proc. 8th IEEE Signal Proc. Workshop Statistical Signal Array Proc.*, pp. 312-315, Corfu, Greece, June 24-26, 1996.

[8] U. Lindgren, T. Wigren and H. Broman, "On local convergence of a class of blind separation algorithms," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 3054-3058, Dec. 1995.

[9] J.K. Tugnait, "Identification of linear stochastic systems via second- and fourth- order cumulant matching," *IEEE Trans. Information Theory*, vol. IT-33, pp. 393-407, May 1987.

[10] H.V. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag: New York, 1988.

[11] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures," in *Proc. 1996 ICASSP*, pp. 2940-2943, Atlanta, GA, May 7-10, 1996.

[12] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input-multi-output systems driven by colored inputs," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 1543-1552, June 1997.

[13] J.K. Tugnait, "Blind estimation and equalization of digital communication FIR channels using cumulant matching," *IEEE Trans. Commun.*, vol. COM-43, pp. 1240-1245, Part III, Feb/March/April 1995.

[14] J.E. Dennis, Jr., D.M. Gray and R.E. Welsch, "Algorithm 573 NL2SOL – An adaptive nonlinear least-squares algorithm [E4]," *ACM Trans. Math. Software*, vol. 7, pp. 369-383, Sept. 1981.

Table 1: *Source-iterative blind signal separation.*

(i)	Maximize (2-2) w.r.t. the equalizer $\mathcal{C}(z)$ to obtain (2-3). Let $\gamma_{4j_0} = \mathrm{CUM}_4(e(k)) = \mathrm{CUM}_4(dw_{j_0}(k))$.
(ii)	Cross-correlate $\{e(k)\}$ (of (2-3)) with the given data (1-1) and define a possibly scaled and shifted estimate of $f_{ij_0}(\tau)$ as $$\widehat{f}_{ij_0}(\tau) := E\{y_i(k)e^*(k-\tau)\} \,/\, E\{\lvert e(k)\rvert^2\}. \qquad \text{(T-1)}$$ Consider now the reconstructed contribution of $e(k)$ to the data $y_i(k)$ ($i = 1, 2, \cdots, M$), denoted by $\widetilde{y}_{i,j_0}(k)$: $$\widetilde{y}_{i,j_0}(k) := \sum_l \widehat{f}_{ij_0}(l)e(k-l). \qquad \text{(T-2)}$$
(iii)	Remove the above contribution from the data to define the outputs of a MIMO system with N outputs and $M-1$ inputs. These are given by $$y_i'(k) := y_i(k) - \widetilde{y}_{i,j_0}(k). \qquad \text{(T-3)}$$
(iv)	If $M > 1$, set $M \leftarrow M-1$, $y_i(k) \leftarrow y_i'(k)$, and go back to **Step (i)**, else quit.

FIGURE CAPTIONS

Fig. 1. Example 1. Amplitude spectra $20\log_{10}\lvert F_{ij}(e^{j\omega})\rvert$ of various subchannels. [Subchannels F_{31} and F_{32} are not shown as $F_{31}(e^{j\omega}) = F_{32}(e^{j\omega}) = 0\ \forall\omega$.]

Fig. 2. Example 1. Normalized mean-square error (T-6) in estimating channel matrix impulse response using various approaches, averaged over 50 Monte Carlo runs. T = record length.

Fig. 3. Example 1. Average SINR (signal-to-interference-and-noise ratio) after blind signal separation using various approaches, averaged over 50 Monte Carlo runs. Record length $T = 1500$.

Fig. 4. Example 1. Average SINR (signal-to-interference-and-noise ratio) after blind signal separation using various approaches, averaged over 50 Monte Carlo runs. Record length $T = 9000$.

Fig. 5. Example 2. Amplitude spectra $20\log_{10}\lvert F_{ij}(e^{j\omega})\rvert$ of various subchannels.

Fig. 6. Example 2. Normalized mean-square error (T-9) in estimating channel matrix impulse response using various approaches, averaged over 50 Monte Carlo runs. T = record length.

Fig. 7. Example 2. Average SINR (signal-to-interference-and-noise ratio) after blind signal separation using various approaches, averaged over 50 Monte Carlo runs. Record length $T = 1500$.

Fig. 8. Example 2. Average SINR (signal-to-interference-and-noise ratio) after blind signal separation using various approaches, averaged over 50 Monte Carlo runs. Record length $T = 9000$.

Table 2: *Performance measures.*

EXAMPLE 1

Normalization: First remove the ambiguities associated with the estimated channel IR (cf. (2-5)). True model (6-1) is such that

$$\sum_{i=1}^{3}\sum_{k=0}^{2}|f_{ij}(k)|^2 = 3 \quad \text{for } j=1 \text{ and } j=2. \tag{T-4}$$

Truncate the estimated IR to 4 samples after alignment with the true IR and then normalize it to satisfy

$$\sum_{i=1}^{3}\sum_{k=-1}^{2}|\widehat{f}_{ij}(k)|^2 = 3 \quad \text{for } j=1 \text{ and } j=2. \tag{T-5}$$

NMSE: The NMSE (normalized mean-square error) is defined as

$$\text{NMSE} = \frac{M_c^{-1}\sum_{l=1}^{M_c}\left[\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{\tau=-1}^{2}\left(\widehat{f}_{ij}^{(l)}(\tau)-f_{ij}(\tau)\right)^2\right]}{\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{\tau=-1}^{2}\left(f_{ij}(\tau)\right)^2} \tag{T-6}$$

where $\widehat{f}_{ij}^{(l)}$ denote the estimate of the ij-th subchannel IR for the l-th Monte Carlo run and there are M_c runs.

SINR: For signal separation the performance measure was taken to be the signal-to-interference-and-noise ratio (SINR) per source signal, defined as

$$\text{SINR}_j = \frac{E\{\|\tilde{\mathbf{y}}^{(j)}(k)\|^2}{E\{\|\tilde{\mathbf{y}}^{(j)}(k)-\widehat{\alpha}\widehat{\tilde{\mathbf{y}}}^{(j)}(k)\|^2\}} \tag{T-7}$$

where $\widehat{\alpha}$ is that value of the scalar α which minimizes $E\{\|\tilde{\mathbf{y}}^{(j)}(k) - \alpha\widehat{\tilde{\mathbf{y}}}^{(j)}(k)\|^2\}$; this is needed to remove the scale ambiguity in the design of (5-3) – it doesn't affect the SINR.

EXAMPLE 2

Normalization: The counterpart to (T-5) is taken as

$$\sum_{i=1}^{3}\sum_{k=-1}^{6}|\widehat{f}_{ij}(k)|^2 = 3 \quad \text{for } j=1 \text{ and } j=2, \tag{T-8}$$

NMSE: The NMSE is modified as

$$\text{NMSE} = \frac{M_c^{-1}\sum_{l=1}^{M_c}\left[\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{\tau=-1}^{2}\left(\widehat{f}_{ij}^{(l)}(\tau)-f_{ij}(\tau)\right)^2\right]}{\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{\tau=-1}^{2}\left(f_{ij}(\tau)\right)^2} \tag{T-9}$$

SINR: As for Example 1.

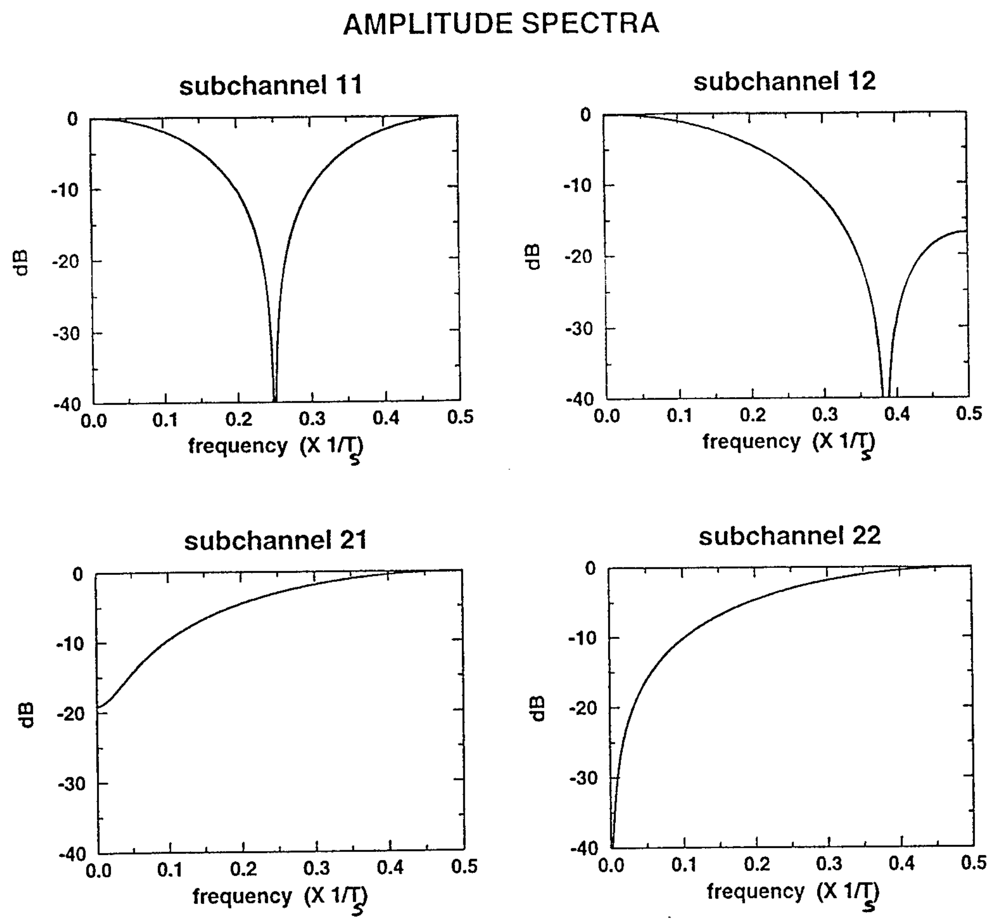

AMPLITUDE SPECTRA
subchannel 11
subchannel 12
subchannel 21
subchannel 22
dB
0
-10
-20
-30
-40
0.0
0.1
0.2
0.3
0.4
0.5
frequency (X 1/T_s)

FIG. 1

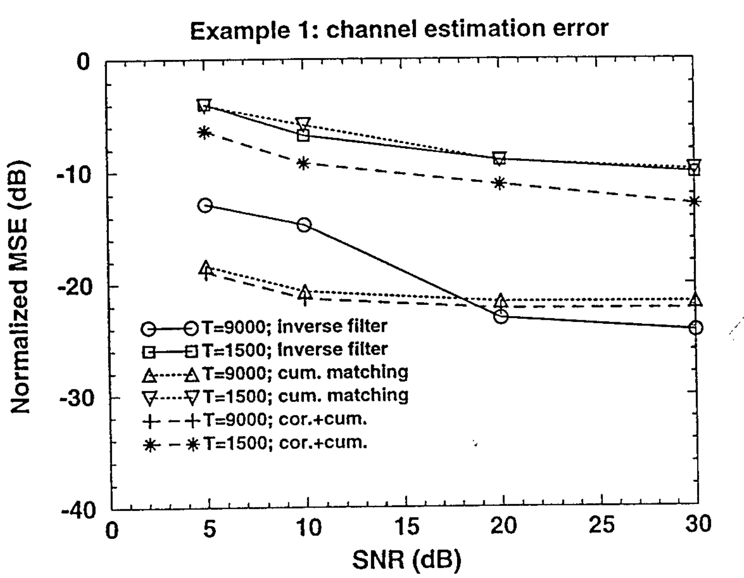
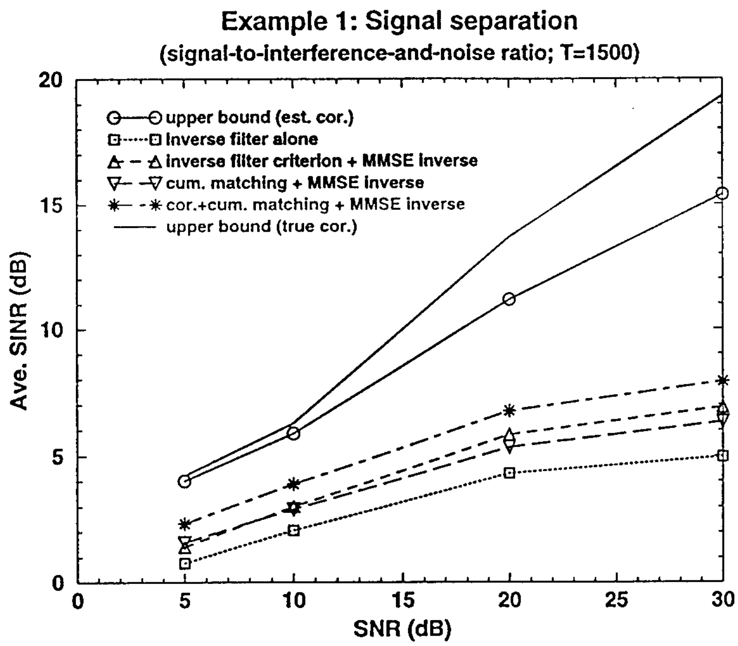
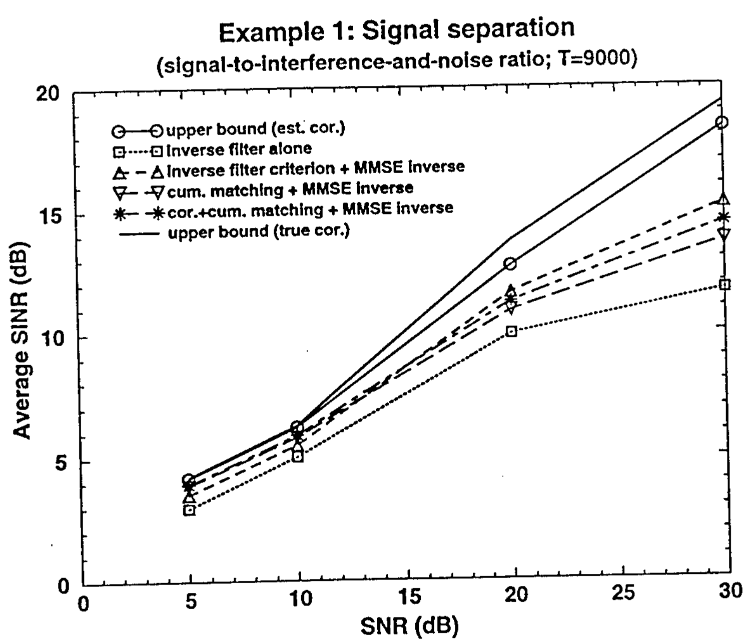

Example 1: channel estimation error
Normalized MSE (dB)
0
-10
-20
-30
-40
0
5
10
15
20
25
30
SNR (dB)
T=9000; inverse filter
T=1500; inverse filter
T=9000; cum. matching
T=1500; cum. matching
T=9000; cor.+cum.
T=1500; cor.+cum.

FIG. 2.



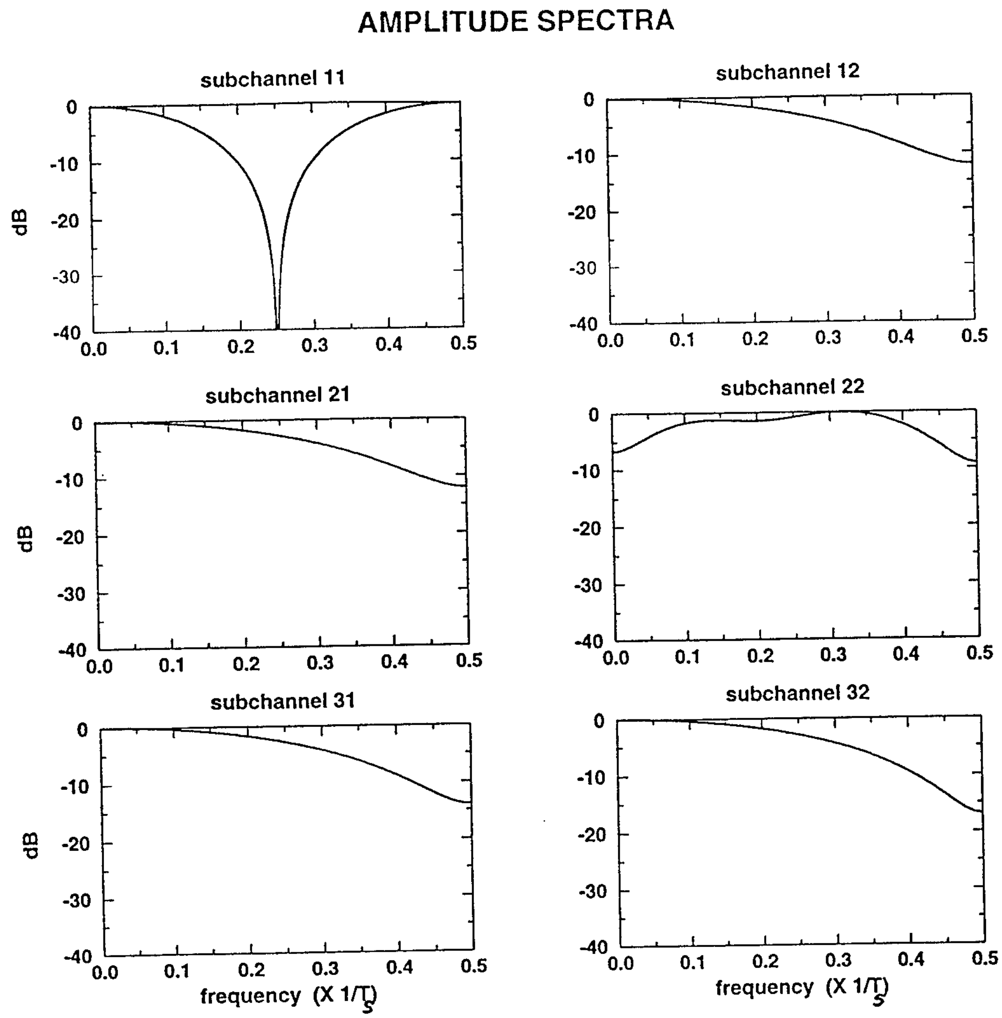
Example 1: Signal separation
(signal-to-interference-and-noise ratio; T=1500)
upper bound (est. cor.)
inverse filter alone
inverse filter criterion + MMSE inverse
cum. matching + MMSE inverse
cor.+cum. matching + MMSE inverse
upper bound (true cor.)
Ave. SINR (dB)
SNR (dB)
0
5
10
15
20
25
30

FIG. 3.



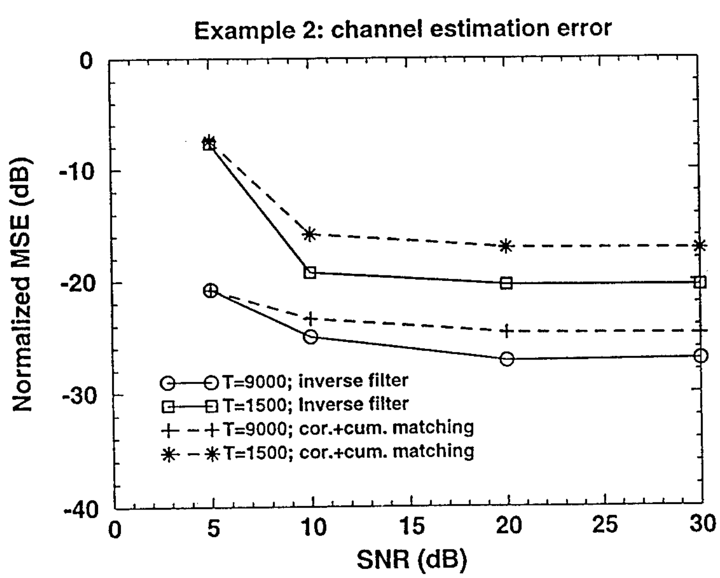
Example 1: Signal separation
(signal-to-interference-and-noise ratio; T=9000)
upper bound (est. cor.)
inverse filter alone
inverse filter criterion + MMSE inverse
cum. matching + MMSE inverse
cor.+cum. matching + MMSE inverse
upper bound (true cor.)
Average SINR (dB)
SNR (dB)
0
5
10
15
20
25
30

FIG. 4.



AMPLITUDE SPECTRA
subchannel 11
subchannel 12
subchannel 21
subchannel 22
subchannel 31
subchannel 32
dB
0
-10
-20
-30
-40
0.0
0.1
0.2
0.3
0.4
0.5
frequency (X 1/T$_S$)

FIG. 5.



Example 2: channel estimation error
Normalized MSE (dB)
0
-10
-20
-30
-40
0
5
10
15
20
25
30
SNR (dB)
T=9000; inverse filter
T=1500; Inverse filter
T=9000; cor.+cum. matching
T=1500; cor.+cum. matching

FIG. 6

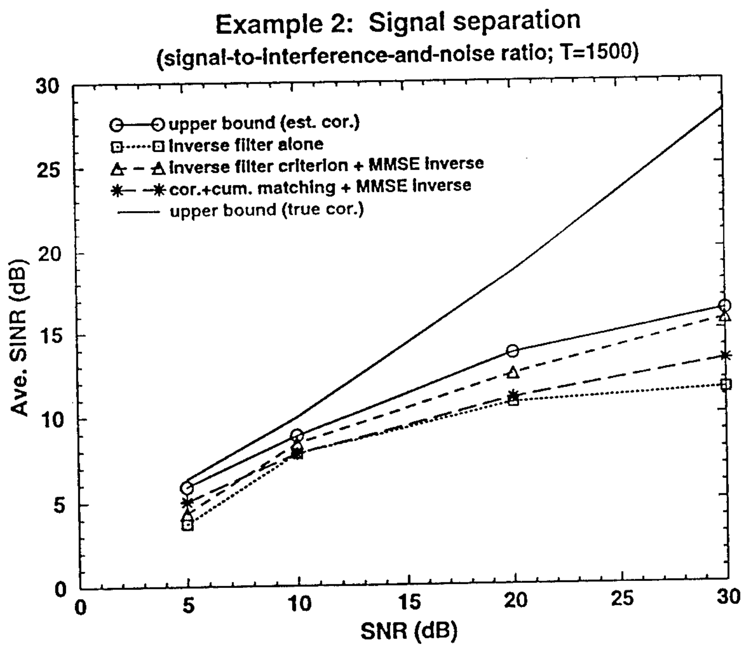

Example 2: Signal separation
(signal-to-interference-and-noise ratio; T=1500)
upper bound (est. cor.)
inverse filter alone
inverse filter criterion + MMSE inverse
cor.+cum. matching + MMSE inverse
upper bound (true cor.)
Ave. SINR (dB)
SNR (dB)
0
5
10
15
20
25
30

FIG. 7

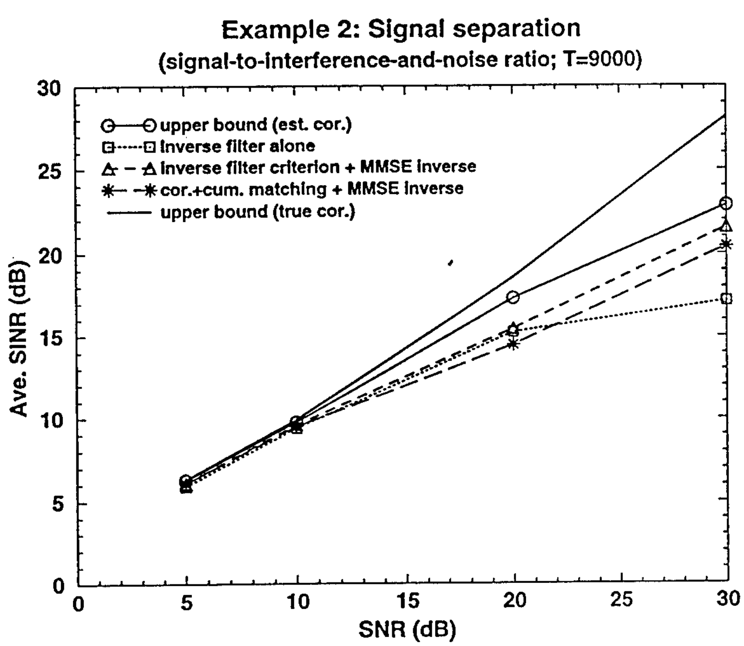

Example 2: Signal separation
(signal-to-interference-and-noise ratio; T=9000)
upper bound (est. cor.)
inverse filter alone
inverse filter criterion + MMSE inverse
cor.+cum. matching + MMSE inverse
upper bound (true cor.)
Ave. SINR (dB)
SNR (dB)
0
5
10
15
20
25
30

FIG. 8

[2]

An Interacting Multiple Model Fixed-Lag Smoothing Algorithm For Markovian Switching Systems[1]

Bing Chen & *Jitendra K. Tugnait*

Department of Electrical Engineering

Auburn University, Auburn, AL 36849, USA

Tel.: (334)844-1846 FAX: (334)844-1809

Email: chenbi1@eng.auburn.edu tugnait@eng.auburn.edu

Abstract

We investigate a suboptimal approach to the fixed-lag smoothing problem for Markovian switching systems. A fixed-lag smoothing algorithm is developed by applying the basic Interacting Multiple Model (IMM) approach to a state-augmented system. The computational load is roughly d (the fixed lag) times beyond that of filtering for the original system. In addition, an algorithm that approximates the "fixed-lag" mode probabilities given measurements up to current time is proposed. The algorithm is illustrated via a target tracking simulation example where a significant improvement over the filtering algorithm is achieved. The IMM fixed-lag smoothing performance for the given example is comparable to that of an existing IMM fixed-interval smoother. Compared to fixed-interval smoothers, the fixed-lag smoothers can be implemented in real-time with a small delay.

[1]This work was supported by the Office of Naval Research under Grant N00014-97-1-0822.

1 Introduction

The system with Markovian switching coefficients considered in this paper is represented by multiple linear models with a given probability of switching between the models. The models are one of the n hypothesized models, $M^1, \cdots, M^n$ for the system, and the event that model j is in effect during the sampling period ending at time t_k (i.e., the sampling period $(t_{k-1}, t_k]$) will be denoted by M_k^j. The state dynamics and measurements, respectively, are modeled as

$$x_k = F_{k-1}^j x_{k-1} + G_{k-1}^j v_{k-1}^j \tag{1}$$

and

$$z_k = H_k^j x_k + w_k^j \tag{2}$$

where x_k is the system state at t_k and of dimension m_x, z_k is the measurement vector at t_k and of dimension m_z, and F_{k-1}^j, G_{k-1}^j, and H_k^j are the system matrices when model j is in effect over the sampling period ending at t_k. The process noise v_{k-1}^j and measurement noise w_k^j are mutually uncorrelated zero-mean white Gaussian processes with covariance matrices Q_{k-1}^j and R_k^j, respectively. At the initial time t_0, the initial conditions for the system state under each model j are assumed to be Gaussian random variables with mean $\overline{x}_0^j$ and covariance P_0^j. The prior statistics $\overline{x}_0^j$ and P_0^j are assumed known, as is $\mu_0^j = P\{M_0^j\}$, the probability of model j at the initial time t_0. The switching from model M_{k-1}^i to model M_k^j is governed by a finite-state stationary Markov chain with transition probabilities $p_{ij} = P\{M_k^j | M_{k-1}^i\}$ which are assumed known.

Motivation for considering system models with switching coefficients (also called stochastic hybrid systems [11]) stems from applicability of such models to a large class of real-world problems such as systems subject to failures/repairs, approximation of nonlinear systems with a set of piecewise linearized models, target tracking, etc. [1], [4], [9]-[11].

This paper is concerned with the problem of state estimation for stochastic hybrid system (1)-(2). This problem has attracted considerable attention in the literature; see [1], [2], [4]-[7], [10], [11] and references therein. Most of the attention has been focused on the filtering problem where one is interested in estimating the state x_k at time k given the current and past measurements $Z_1^k = \{z_1, z_2, \cdots, z_k\}$. The optimal MMSE (minimum mean-square error) filter requires n^k Kalman filters in parallel in order to obtain the optimal state filtered-estimate at time k. Thus the optimal approach is not practical and suboptimal techniques have to be considered. Several suboptimal techniques have been investigated in the literature [1], [2], [11]. The interacting multiple model (IMM) algorithm of [2] has been found to offer a good compromise between the computational and storage requirements and estimation accuracy [10],[11].

The state smoothing problem for stochastic hybrid systems has attracted much less attention. Here one is interested in estimating the state x_k given past and future data Z_1^N ($N \geq k$). Fixed-interval smoothing problem (where record length N is fixed) has been considered in [6] and [7]. Both [6] and [7] have used some versions of the IMM algorithm in order to implement fixed-interval smoothing. It is stated in [6, Sec. VIII] that "... we believe that the time-reversion and smoothing techniques developed are of interest to other hybrid state estimation problems ... This leads immediately to the question if and how IMM-smoothing can be extended to fixed-lag smoothing." This paper is concerned with the problem of fixed-lag smoothing using an IMM approach. In fixed-lag smoothing with lag d ($d \geq 0$) one is interested in estimating the state x_k given past and part of future data Z_1^{k+d} where d is fixed. Equivalently, one looks for estimate of x_{k-d} given data Z_1^k. For $d = 0$ we have the filtering solution.

For linear systems with completely known parameters, it is well known that fixed-lag smoothing leads to an improvement in the performance (at the cost of increased computational complexity) when compared with the zero-lag case (filtering) [8]. Indeed, in most

cases, a 'small' lag leads to a performance almost as good as that due to fixed-interval smoothing [8]. An advantage of fixed-lag smoothing over fixed-interval smoothing is that the former can be implemented in real time with a small fixed time-delay whereas the latter has to wait for the entire measurement record.

Fixed-lag smoothing for stochastic hybrid systems has been investigated in [5]. In [5] a hypothesis-pruning approach (called detection-estimation [1]) has been considered for state estimation. Since the IMM algorithm (which belongs to the class of generalized pseudo-Bayes algorithms [2]) has been found to perform better than the hypothesis-pruning approaches for the same computational complexity, it is of some interest to investigate IMM algorithm based fixed-lag smoothing.

The paper is organized as follows. The basic IMM filtering algorithm is reviewed in Sec. 2. A state-augmentation approach is followed in Sec. 3 to derive an IMM fixed-lag smoothing algorithm via the IMM filtering algorithm discussed in Sec. 2. Sec. 3 is focused on state estimation. It is of considerable interest to compute the conditional mode probabilities $P(M_k^j|Z_1^{k+d})$ (given the data). Certain approximations are suggested in Sec. 4 to compute these probabilities. A discussion of the computational requirements of the proposed fixed-lag IMM smoother as compared with that of the IMM filter is provided in Sec. 5. In Sec. 6 we illustrate the proposed approach via a target tracking simulation example taken from [7]. When compared with the results of [7], it shows that a delay of just a few samples leads to a performance comparable to that of the fixed-interval smoothing of [7]. Finally, some concluding remarks are provided in Sec. 7.

2 Basic IMM Algorithm

The IMM algorithm [2] for state filtering is based on running n "mode-matched" state estimation filters which exchange information (interact) at each sampling instant. It assumes that

the conditional probability density $f(x_k|M_k^j, Z_1^k)$ is Gaussian with mean $\widehat{x}_{k|k}^j = E\{x_k|M_k^j, Z_1^k\}$ and covariance $P_{k|k}^j = E\{[x_k - \widehat{x}_{k|k}^j][x_k - \widehat{x}_{k|k}^j]'|M_k^j, Z_1^k\}$ where the symbol $'$ denotes the transpose operation. In reality, however, the density $f(x_k|M_k^j, Z_1^k)$ is a Gaussian sum (containing n^k terms).

As the algorithm is well-explained in [1] (see also [10] and [11]), we will only briefly outline below the basic steps in "one cycle" (i.e. processing needed to update for a new measurement) of the IMM filtering algorithm. We follow Table I of [10] for most part.

Initialization: Given $\widehat{x}_{k-1|k-1}^j$, the associated covariance matrix $P_{k-1|k-1}^j$ and the conditional mode probability $\mu_{k-1}^j := P(M_{k-1}^j|Z_1^{k-1})$ for each $j \in \mathcal{M}_n := \{1, 2, \cdots, n\}$. For $k = 1$, we take $\widehat{x}_{0|0}^j = \overline{x}_0^j$, $P_{0|0}^j = P_0^j$ and $\mu_0^j = P(M_0^j)$.

Interaction ($\forall j \in \mathcal{M}_n$):

predicted mode probability:

$$\mu^{j-} := P\{M_k^j|Z_1^{k-1}\} = \sum_{i=1}^{n} p_{ij}\mu_{k-1}^i \tag{3}$$

mixing probability:

$$\mu^{i|j} := P\{M_{k-1}^i|M_k^j, Z_1^{k-1}\} = p_{ij}\mu_{k-1}^i/\mu^{j-} \tag{4}$$

mixed estimate:

$$\widehat{x}_{k-1|k-1}^{0j} := E[x_{k-1}|M_k^j, Z_1^{k-1}] = \sum_{i=1}^{n} \widehat{x}_{k-1|k-1}^i \mu^{i|j} \tag{5}$$

covariance of the mixed estimate:

$$\begin{aligned} P_{k-1|k-1}^{0j} &= E\{[x_k - \widehat{x}_{k-1|k-1}^{0j}][x_k - \widehat{x}_{k-1|k-1}^{0j}]'|M_k^j, Z_1^{k-1}\} \\ &= \sum_{i=1}^{n} \left\{ P_{k-1|k-1}^i + [\widehat{x}_{k-1|k-1}^i - \widehat{x}_{k-1|k-1}^{0j}][\widehat{x}_{k-1|k-1}^i - \widehat{x}_{k-1|k-1}^{0j}]' \right\} \mu^{i|j} \end{aligned} \tag{6}$$

Prediction and filtering ($\forall j \in \mathcal{M}_n$):

$$\widehat{x}_{k|k-1}^j = E\{x_k|M_k^j, Z_1^{k-1}\} = F_{k-1}^j \widehat{x}_{k-1|k-1}^{0j} \tag{7}$$

$$
\begin{aligned}
P^j_{k|k-1} &= E\{[x_k - \widehat{x}^j_{k|k-1}][x_k - \widehat{x}^j_{k|k-1}]' | M^j_k, Z^{k-1}_1\} \\
&= F^j_{k-1} P^{0j}_{k-1|k-1} {F^j_{k-1}}' + G^j_{k-1} Q^j_{k-1} {G^j_{k-1}}'
\end{aligned}
\tag{8}
$$

measurement residual:

$$
\nu^j_k := z_k - H^j_k \widehat{x}^j_{k|k-1} \tag{9}
$$

residual covariance:

$$
S^j_k := E\{\nu^j_k \nu^{j'}_k\} = H^j_k P^j_{k|k-1} H^{j'}_k + R^j_k \tag{10}
$$

filter gain:

$$
W^j_k = P^j_{k|k-1} H^{j'}_k S^{j^{-1}}_k \tag{11}
$$

filtered state estimate:

$$
\widehat{x}^j_{k|k} = E\{x_k | M^j_k, Z^k_1\} = \widehat{x}^j_{k|k-1} + W^j_k \nu^j_k \tag{12}
$$

covariance of the filtered state estimate:

$$
\begin{aligned}
P^j_{k|k} &= E\{[x_k - \widehat{x}^j_{k|k}][x_k - \widehat{x}^j_{k|k}]' | M^j_k, Z^k_1\} \\
&= P^j_{k|k-1} - W^j_k S^j_k W^{j'}_k
\end{aligned}
\tag{13}
$$

likelihood function:

$$
\Lambda^j_k = \mathcal{N}(\nu^j_k; 0, S^j_k) := |2\pi S^j_k|^{-1/2} \exp\left[-\frac{1}{2} \nu^{j'}_k (S^j_k)^{-1} \nu^j_k\right] \tag{14}
$$

mode probability:

$$
\mu^j_k = P(M^j_k | Z^k_1) = \frac{\mu^{j-} \Lambda^j_k}{\sum_{i=1}^n \mu^{i-} \Lambda^i_k} \tag{15}
$$

Combination:

$$
\widehat{x}_{k|k} = E[x_k | Z^k_1] = \sum_{i=1}^n \widehat{x}^i_{k|k} \mu^i_k \tag{16}
$$

$$
\begin{aligned}
P_{k|k} &= E\{[x_k - \widehat{x}_{k|k}][x_k - \widehat{x}_{k|k}]' | Z^k_1\} \\
&= \sum_{i=1}^n \left\{ P^i_{k|k} + [\widehat{x}^i_{k|k} - \widehat{x}_{k|k}][\widehat{x}^i_{k|k} - \widehat{x}_{k|k}]' \right\} \mu^i_k
\end{aligned}
\tag{17}
$$

3 IMM Fixed-Lag Smoothing Algorithm

For some fixed-lag d and all $k \geq d$, our objective is to find the fixed-lag smoothing state estimate

$$\widehat{x}_{k-d|k} = E[x_{k-d}|Z_1^k] \tag{18}$$

and the associated error covariance matrix

$$P_{k-d|k} = E\{ [x_{k-d} - \widehat{x}_{k-d|k}][x_{k-d} - \widehat{x}_{k-d|k}]'|Z_1^k \}. \tag{19}$$

When $d = 0$, we have the filtered state estimate as discussed in Sec. 2. We will follow a state-augmentation approach to define a larger dynamical stochastic hybrid system and then apply the results of Sec. 2 to this augmented system. State augmentation for derivation of fixed-lag smoothing estimators has been used before for "non-switching" linear systems [8, Sec. 7.3] but not for stochastic hybrid systems.

Augment the state variable x_k to $\widetilde{x}_k$ as

$$\widetilde{x}_k' = [\widetilde{x}_k^{(0)'}, \;\; x_k^{(1)'}, \;\; \cdots, \;\; x_k^{(d)'}] \tag{20}$$

where

$$\widetilde{x}_k^{(0)} = x_k, \;\; \widetilde{x}_k^{(1)} = x_{k-1}, \;\; \cdots, \;\; x_k^{(d)} = x_{k-d}. \tag{21}$$

Suppose that for the augmented system, we obtain the filtered state estimate

$$\widehat{\widetilde{x}}_{k|k} := E\{\widetilde{x}_k|Z_1^k\} \tag{22}$$

and the associated covariance matrix

$$\widetilde{P}_{k|k} := E\{[\widetilde{x}_k - \widehat{\widetilde{x}}_{k|k}][\widetilde{x}_k - \widehat{\widetilde{x}}_{k|k}]'|Z_1^k\}. \tag{23}$$

It therefore follows that

$$\widehat{\widetilde{x}}_{k|k}^{(i)} := E[\widetilde{x}_k^{(i)}|Z_1^k] = \widehat{x}_{k-i|k} \tag{24}$$

and

$$\widetilde{P}^{(i,i)}_{k|k} := E\{[\widetilde{x}^{(i)}_k - \widehat{\widetilde{x}}^{(i)}_{k|k}][\widetilde{x}^{(i)}_k - \widehat{\widetilde{x}}^{(i)}_{k|k}]'|Z^k_1\} = P_{k-i|k} \tag{25}$$

for $i = 0, \cdots, d$, where

$$\widetilde{P}_{k|k} = \begin{bmatrix} \widetilde{P}^{(0,0)}_{k|k} & \widetilde{P}^{(0,1)}_{k|k} & \cdots & \widetilde{P}^{(0,d)}_{k|k} \\ \widetilde{P}^{(1,0)}_{k|k} & \widetilde{P}^{(1,1)}_{k|k} & \cdots & \widetilde{P}^{(1,d)}_{k|k} \\ \vdots & \vdots & & \vdots \\ \widetilde{P}^{(N,0)}_{k|k} & \widetilde{P}^{(N,1)}_{k|k} & \cdots & \widetilde{P}^{(d,d)}_{k|k} \end{bmatrix}.$$

Note that $\widetilde{P}_{k|k}$ is symmetric, i.e., $\widetilde{P}^{(i,j)}_{k|k} = \widetilde{P}^{(j,i)\prime}_{k|k}$.

Using (1), (2), (20) and (21), the augmented system can be written as follows:

$$\begin{bmatrix} \widetilde{x}^{(0)}_k \\ \widetilde{x}^{(1)}_k \\ \widetilde{x}^{(2)}_k \\ \vdots \\ \widetilde{x}^{(d)}_k \end{bmatrix} = \begin{bmatrix} F^j_{k-1} & 0 & \cdots & 0 & 0 \\ I & 0 & & 0 & 0 \\ 0 & I & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \widetilde{x}^{(0)}_{k-1} \\ \widetilde{x}^{(1)}_{k-1} \\ \widetilde{x}^{(2)}_{k-1} \\ \vdots \\ \widetilde{x}^{(d)}_{k-1} \end{bmatrix} + \begin{bmatrix} G^j_{k-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} v^j_{k-1} \tag{26}$$

and

$$z_k = [H^j_k \;\; 0 \;\; \cdots \;\; 0 \;\; 0] \begin{bmatrix} \widetilde{x}^{(0)}_k \\ \widetilde{x}^{(1)}_k \\ \widetilde{x}^{(2)}_k \\ \vdots \\ \widetilde{x}^{(d)}_k \end{bmatrix} + w^j_k. \tag{27}$$

The above augmented state and measurement equations may be written more compactly as

$$\widetilde{x}_k = \widetilde{F}^j_{k-1}\widetilde{x}_{k-1} + \widetilde{G}^j_{k-1}v^j_{k-1} \tag{28}$$

and

$$z_k = \widetilde{H}^j_k\widetilde{x}_k + w^j_k \tag{29}$$

where the system matrices $\widetilde{F}^j_{k-1}$, $\widetilde{G}^j_{k-1}$, and $\widetilde{H}^j_k$ are defined in an obvious manner.

We now apply the basic IMM algorithm to the augmented system. Unlike Sec. 2 where $f(x_k|M^j_k, Z^k_1)$ is approximated by a Gaussian random vector, now we approximate $f(\widetilde{x}_k|M^j_k, Z^k_1)$ $= f(x_k, x_{k-1}, \cdots, x_{k-d}|M^j_k, Z^k_1)$ by a Gaussian random vector. Clearly the latter approximation implies the former whereas the converse is not true in general. The resulting algorithm is as follows:

Interaction ($\forall j \in \mathcal{M}_n$):

predicted mode probability:

$$\mu^{j-} := P\{M^j_k|Z^{k-1}_1\} = \sum_{i=1}^{n} p_{ij}\mu^i_{k-1} \tag{30}$$

mixing probability:

$$\mu^{i|j} := P\{M^i_{k-1}|M^j_k, Z^{k-1}_1\} = p_{ij}\mu^i_{k-1}/\mu^{j-} \tag{31}$$

mixed estimate:

$$\widehat{\widetilde{x}}^{0j}_{k-1|k-1} = E[\widetilde{x}_{k-1}|M^j_k, Z^{k-1}_1] = \sum_{i=1}^{n} \widehat{\widetilde{x}}^i_{k-1|k-1}\mu^{i|j} \tag{32}$$

covariance of the mixed estimate:

$$\begin{aligned}\widetilde{P}^{0j}_{k-1|k-1} &= E\{[\widetilde{x}_k - \widehat{\widetilde{x}}^{0j}_{k-1|k-1}][\widetilde{x}_k - \widehat{\widetilde{x}}^{0j}_{k-1|k-1}]'|M^j_k, Z^{k-1}_1\} \\ &= \sum_{i=1}^{n} \left\{\widetilde{P}^i_{k-1|k-1} + [\widehat{\widetilde{x}}^i_{k-1|k-1} - \widehat{\widetilde{x}}^{0j}_{k-1|k-1}][\widehat{\widetilde{x}}^i_{k-1|k-1} - \widehat{\widetilde{x}}^{0j}_{k-1|k-1}]'\right\}\mu^{i|j}.\end{aligned} \tag{33}$$

We will only be interested in $\widetilde{P}^{0j(i,i)}_{k-1|k-1}$, the diagonal sub-matrices of $\widetilde{P}^{0j}_{k-1|k-1}$, and in $\widetilde{P}^{0j(0,i)}_{k-1|k-1}$, for $i = 0, 1, \cdots, d-1$, in order to complete the filtering process in the sequel [8, Sec. 7.3]; see also (37) later in this paper. Thus in (33) we only need to compute $\widetilde{P}^{0j(i,i)}_{k-1|k-1}$ and $\widetilde{P}^{0j(0,i)}_{k-1|k-1}$ for $i = 0, \cdots, d-1$.

Prediction and filtering ($\forall j \in \mathcal{M}_n$):

Using (7), (8) and (26), it follows that

$$\hat{\tilde{x}}^{j(0)}_{k|k-1} = E\{\tilde{x}^{(0)}_k | M^j_k, Z^{k-1}_1\} = F^j_{k-1}\hat{\tilde{x}}^{0j(0)}_{k-1|k-1} \tag{34}$$

$$\hat{\tilde{x}}^{j(i)}_{k|k-1} = E\{\tilde{x}^{(i)}_k | M^j_k, Z^{k-1}_1\} = \hat{\tilde{x}}^{0j(i-1)}_{k-1|k-1} \text{ for } i = 1, \cdots, d \tag{35}$$

$$\tilde{P}^{j(0,0)}_{k|k-1} = E\{[\tilde{x}^{(i)}_k - \hat{\tilde{x}}^{j(i)}_{k|k-1}][\tilde{x}^{(i)}_k - \hat{\tilde{x}}^{j(i)}_{k|k-1}]' | M^j_k, Z^{k-1}_1\}$$

$$= F^j_{k-1}\tilde{P}^{0j(0,0)}_{k-1|k-1}{F^j_{k-1}}' + G^j_{k-1}Q^j_{k-1}{G^j_{k-1}}' \tag{36}$$

$$\tilde{P}^{j(i,i)}_{k|k-1} = \tilde{P}^{0j(i-1,i-1)}_{k|k-1} \text{ for } i = 1, \cdots, d \tag{37}$$

and

$$\tilde{P}^{j(0,i)}_{k|k-1} = F^j_{k-1}\tilde{P}^{0j(0,i)}_{k-1|k-1} \text{ for } i = 1, \cdots, d. \tag{38}$$

Using (9), (10) and (27), it follows that

measurement residual:

$$\nu^j_k := z_k - H^j_k\hat{\tilde{x}}^{j(0)}_{k|k-1} \tag{39}$$

residual covariance:

$$S^j_k := E\{\nu^j_k\nu^{j'}_k\} = H^j_k\tilde{P}^{j(0,0)}_{k|k-1}H^{j'}_k + R^j_k. \tag{40}$$

filter gain: Using (11) it follows that

$$W^{j(i)}_k = \tilde{P}^{j(0,i)'}_{k|k-1}H^{j'}_k S^{j^{-1}}_k \text{ for } i = 0, \cdots, d. \tag{41}$$

Using (12) and (13) we have

$$\hat{\tilde{x}}^{j(i)}_{k|k} = \hat{\tilde{x}}^{j(i)}_{k|k-1} + W^{j(i)}_k\nu^j_k \tag{42}$$

$$\tilde{P}^{j(i,i)}_{k|k} = \tilde{P}^{j(i,i)}_{k|k-1} - W^{j(i)}_k S^j_k W^{j(i)'}_k \text{ for } i = 0, 1, \cdots, d \tag{43}$$

$$\tilde{P}^{j(0,i)}_{k|k} = \tilde{P}^{j(0,i)}_{k|k-1} - W^{j(0)}_k S^j_k W^{j(i)'}_k \text{ for } i = 1, \cdots, d. \tag{44}$$

likelihood function:

$$\Lambda_k^j = \mathcal{N}(\nu_k^j; 0, S_k^j) = |2\pi S_k^j|^{-1/2} \exp\left[-\frac{1}{2}\nu_k^{j\prime}(S_k^j)^{-1}\nu_k^j\right] \tag{45}$$

mode probability:

$$\mu_k^j = P(M_k^j|Z_1^k) = \frac{\mu^{j-}\Lambda_k^j}{\sum_{i=1}^n \mu^{i-}\Lambda_k^i} \tag{46}$$

Combination:

$$\hat{\tilde{x}}_{k|k} := E[\tilde{x}_k|Z_1^k] = \sum_{i=1}^n \hat{\tilde{x}}_{k|k}^i \mu^i \tag{47}$$

$$\begin{aligned}\tilde{P}_{k|k} &= E\{[\tilde{x}_k - \hat{\tilde{x}}_{k|k}][\tilde{x}_k - \hat{\tilde{x}}_{k|k}]'|Z_1^k\} \\ &= \sum_{i=1}^n \left\{\tilde{P}_{k|k}^i + [\hat{\tilde{x}}_{k|k}^i - \hat{\tilde{x}}_{k|k}][\hat{\tilde{x}}_{k|k}^i - \hat{\tilde{x}}_{k|k}]'\right\}\mu_k^i\end{aligned} \tag{48}$$

In (48), as before, we need not compute all the elements as we are only interested in $\tilde{P}_{k|k}^{(0,i)}$ and $\tilde{P}_{k|k}^{(i,i)}$ for $i = 0, \cdots, d$.

Finally we obtain the smoothed state estimates (in addition to the current state estimate):

$$\hat{x}_{k-i|k} = \hat{\tilde{x}}_{k|k}^{(i)} \tag{49}$$

and the associated error covariance matrix

$$P_{k-i|k} = \tilde{P}_{k|k}^{(i,i)} \tag{50}$$

for $i = 0, \cdots, d$.

Initialization for the Augmented System: In order to let the augmented system have the same dynamics as the original system, we set

$$\tilde{x}_0 = [x_0' \ \ 0 \ \ \ldots \ \ 0]' \tag{51}$$

which implies that

$$\hat{\tilde{x}}_{0|0}^{j(0)} = \bar{x}_0^j \quad \text{and} \quad \hat{\tilde{x}}_{0|0}^{j(i)} = 0 \ \text{ for } i \neq 0 \tag{52}$$

and

$$\widetilde{P}^{j(0,0)}_{0|0} = P^j_0 \quad \text{and} \quad \widetilde{P}^{j(k,l)}_{0|0} = 0 \ \text{ for } (k,l) \neq (0,0). \tag{53}$$

Recall that for the original system we have

$$\widehat{x}_{0|0} = \overline{x}^j_0 \quad \text{and} \quad P^j_{0|0} = P^j_0. \tag{54}$$

4 Approximation of Mode Probabilities

In Sec. 3, we obtained only the mode probability at t_k in (46). In keeping with fixed-lag smoothing, we would also like to obtain the mode probabilities $P(M^j_{k-i}|Z^k_1)$ for $i = 1, \cdots, d$. Following some of the approximations made in [7], we make an approximation by replacing Z^k_1 with $\{\widehat{x}_{k-i|k}, Z^{k-i}_1\}$, i.e.,

$$\begin{aligned} P(M^j_{k-i}|Z^k_1) &\approx P(M^j_{k-i}|\widehat{x}_{k-i|k}, Z^{k-i}_1) \\ &= \frac{1}{c}[f(\widehat{x}_{k-i|k}|M^j_{k-i}, Z^{k-i}_1)P(M^j_{k-i}|Z^{k-i}_1)] \end{aligned} \tag{55}$$

where c in (55) is a normalization constant given by

$$c = \sum_{j=1}^{n} f(\widehat{x}_{k-i|k}|M^j_{k-i}, Z^{k-i}_1)P(M^j_{k-i}|Z^{k-i}_1)]. \tag{56}$$

In a manner similar to that in Eqns. (75) and (84) of [7], we have replaced the measurements Z^k_{k-i+1} in (55) with the smoothed state estimate $\widehat{x}_{k-i|k}$. We also make the approximations

$$\begin{aligned} f(\widehat{x}_{k-i|k}|M^j_{k-i},, Z^{k-i}_1) &\approx f(x_{k-i}|M^j_{k-i},, Z^{k-i}_1) \\ &\approx \mathcal{N}(x_{k-i}; \widehat{x}^j_{k-i|k-i}, P^j_{k-i|k-i}) \\ &\approx \mathcal{N}(\widehat{x}_{k-i|k}; \widehat{x}^j_{k-i|k-i}, P^j_{k-i|k-i}). \end{aligned} \tag{57}$$

where

$$\mathcal{N}(x; y, P) := |2\pi P|^{-1/2} \exp\left[-\frac{1}{2}(x-y)'P^{-1}(x-y)\right]. \tag{58}$$

Therefore, (55) can be rewritten as

$$P(M_{k-i}^{j}|Z_1^k) \approx \frac{1}{c'}\mathcal{N}(\widehat{x}_{k-i|k};\widehat{x}_{k-i|k-i}^{j},P_{k-i|k-i}^{j})P(M_{k-i}^{j}|Z_1^{k-i}) \tag{59}$$

where normalization constant c' is given by

$$c' = \sum_{j=1}^{n}\mathcal{N}(\widehat{x}_{k-i|k};\widehat{x}_{k-i|k-i}^{j},P_{k-i|k-i}^{j})P(M_{k-i}^{j}|Z_1^{k-i}).$$

We note that in (59), $P(M_{k-i}^{j}|Z_1^{k-i})$ is the "old" mode probability based on filtered state estimates when measurements only up to t_{k-i} are available. It is the likelihood function that $\mathcal{N}(\widehat{x}_{k-i|k};\widehat{x}_{k-i|k-i}^{j},P_{k-i|k-i}^{j})$ utilizes the new information contained in measurements after time t_{k-i}.

It should be noted that (58) can not be applied if $|P| = 0$. As will be seen in Sec. 6 in a target tracking context, for constant velocity models with accleration as a system state, such a situation can arise. Our solution (discussed in more detail in Sec. 6) is to use a state of reduced order such that the corresponding covariance matrix is of full rank.

5 Analysis of Computational Load

Here we carry out a "crude" comparison of the IMM smoothing algorithm of Secs. 3 and 4 with the original non-augmented system IMM filtering algorithm (see Sec. 2) regarding their relative computational requirements

During interaction, comparing (3) and (4) with (30) and (31), respectively, it is seen that the computational loads are identical. Comparing (5) and (6) with (32) and (33), respectively, it is seen that the latter needs more multiplications: the computational load of (32) is d times that of (5) and the computational load of (33) is originally d^2 times that of (6), but since we only need the diagonal sub-matrices and the sub-matrices in the first column of $\widetilde{P}_{k-1|k-1}^{0j}$, it turns out that the computational load of (33) is reduced to about $2d$ times that of (6).

During prediction and filtering, comparing (7) and (8) with (34)–(38), we first note that (35) and (37) do not need any additional computations at all. Furthermore, (34) and (36) have the same computational load as that for (7) and (8). The only computational increase here for the smoothing algorithm is in (38). There is no relative computational load increase in computing measurement residual (see (9) and (39)), residual covariance (see (10) and (40)), likelihood function (see (14) and (45)) and mode probability (see (15) and (46)). There is indeed some computational load increase when computing the filter gain (see (11) and (41)), but this only involves matrix multiplication (and not other complex operations such as computing inverse of a matrix). The same is true for state estimate and its error covariance matrix for each mode (see (12)–(13) and (42)–(44)).

During combination, the relative computational load increases are similar to that for (32) and (33), i.e., of the order of d times that of (17).

The mode probability $P(M^j_{k-N}|Z^k_1)$ calculations in (59) are of the same order as that for (46) and (15).

Overall we run n filters/smoothers in parallel whereas [7] runs n^2 smoothers in parallel and [6] runs n smoothers in parallel. Unlike [6] and [7], we do not need a backward-time model (and its "initial" conditions at final time). More significantly, we run the smoother (beyond the filtering part) only for d samples whereas [6] and [7] run it over the entire measurement record.

6 Simulation Example

A target tracking example is provided to compare the performance of the IMM fixed-lag smoothing algorithm and the (forward-time) IMM filtering algorithm. This example has been taken from [7]. The following scenario is considered. A target is moving in a two-dimensional plane with a constant speed and performing two constant-speed $3g$ maneuvers. The first maneuver occurs from 10 to 22 s, and the second one from 26 to 38 s. The true position, velocity and acceleration of the target are shown in Fig. 1. Position measurements (range and bearing) of the target are sampled with period $T = 1$ s. The measurements contain zero-mean Gaussian errors with standard deviations of 15 m in range and 0.002 rad in bearing.

The state of the target is defined as

$$x = [\xi \ \ \dot{\xi} \ \ \ddot{\xi} \ \ \eta \ \ \dot{\eta} \ \ \ddot{\eta}] \tag{60}$$

with ξ and η denoting the orthogonal (Cartesian) coordinates of the horizontal plane and $\dot{\xi} := \frac{d\xi}{dt}$. The discrete-time multiple model set consists of two models ($n = 2$ in (1)-(2)) as:

1) The Constant Velocity (CV) model ($j = 1$ in (1)-(2)) with a piecewise-constant acceleration process noise and the noise covariance $Q_{cv} = 0.25 I_2$ m^2/s^4 where I_2 is the 2×2 identity matrix. The corresponding system matrices are [3]:

$$F_k^1 = F_{cv} = \begin{bmatrix} 1 & T & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad G_k^1 = G_{cv} = \begin{bmatrix} \frac{1}{2}T^2 & 0 \\ T & 0 \\ 0 & 0 \\ 0 & \frac{1}{2}T^2 \\ 0 & T \\ 0 & 0 \end{bmatrix}. \tag{61}$$

2) The Constant Acceleration (CA) model with a piecewise-constant jerk process noise and the noise covariance $Q_{ca} = 9I_2 \ \mathrm{m}^2/\mathrm{s}^6$. The corresponding system matrices are:

$$F_k^2 = F_{ca} = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & \frac{1}{2}T^2 \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad G_k^2 = G_{ca} = \begin{bmatrix} \frac{1}{6}T^3 & 0 \\ \frac{1}{2}T^2 & 0 \\ T & 0 \\ 0 & \frac{1}{6}T^3 \\ 0 & \frac{1}{2}T^2 \\ 0 & T \end{bmatrix}. \tag{62}$$

The initial model probabilities are $\mu_0^1 = 0.9$ and $\mu_0^2 = 0.1$ (as in [7]). The model-switching probability matrix is given by

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix}. \tag{63}$$

The initial estimates of the velocity and acceleration are arbitrarily set to zero, with variances of $10^6 \mathrm{m}^2/\mathrm{s}^2$ and $10^6 \mathrm{m}^2/\mathrm{s}^4$, respectively, as in [7].

Let $x_{(i)k}$ denote the i–th element of vector x_k which is a 6-vector (cf. (60)). The measurements are given by

$$z_k = \begin{bmatrix} \sqrt{x_{(1)k}^2 + x_{(4)k}^2} \\ \arctan\left(x_{(4)k}/x_{(1)k}\right) \end{bmatrix} + w_k \tag{64}$$

where the measurement equation is the same for the two models in the model set. The covariance matrix of the 2-vector w_k is given by (as in [7])

$$R_k^1 = R_k^2 = \begin{bmatrix} (15 \ \mathrm{m})^2 & 0 \\ 0 & (0.002 \ \mathrm{rad})^2 \end{bmatrix}. \tag{65}$$

A first-order Taylor series expansion around $\hat{x}_{k|k-1}^j$ was used to linearize (64), i.e. a first-order extended Kalman filter (EKF) was used to apply the various estimation algorithms.

The fixed-lag IMM smoothing algorithm was implemented using a first-order EKF over 100 Monte Carlo runs. Note that due to the fact that all the elements in the 3rd and 6th

rows of G_{cv} are zero, the determinant of $P^1_{k-i|k-i}$ is zero where model $j = 1$ is the CV model. Therefore in this example we can not apply (57) directly. Instead we use a state of reduced order with $x_r := [\xi \;\; \dot{\xi} \;\; \eta \;\; \dot{\eta}]$ so that we can evaluate (57). The order of the corresponding covariance matrix $P^j_{k-i|k-i}$ in (57) is reduced to match the state x_r accordingly. Note we need to use x_r for both the CV model and the CA model, in order to calculate the likelihood function defined in (57). This 'adaptation' is reasonable because we should weight the probabilities of all models based on the same set of states.

Fig. 2 displays the average root-mean-square errors (RMSE) in position, velocity and acceleration, and the average CV model probabilities. The legend used in Fig. 2 is self-explanatory: 0 stands for the case of (forward-time) IMM filtering (no smoothing, or fixed-lag $d = 0$) and 1, 2 and 3 stand for the case of the proposed fixed-lag IMM smoothing algorithm with fixed-lags d =1, 2 and 3, respectively. The thick solid line in Fig. 2(d) stands for the normalized magnitude of the acceleration. It can be seen from Fig. 2(a)-(c) that the various RMSE's using the proposed smoothing algorithm decrease with increasing lag d. Comparing Fig. 2 with Fig. 4 in [7], it is seen that the performance of our algorithm for $d = 3$ almost approaches that of fixed-interval smoothing algorithm presented in [7]. When $d = 1$, the most significant reduction in RMSE occurs where the peak RMSE in position is reduced from 145 m (no smoothing) to 65 m while the peak RMSE in velocity is reduced from 102 m/s (no smoothing) to 78 m/s. Besides, significant reductions in RMSE are also achieved over the entire tracking interval.

Fig. 2(d) displays the average CV model probability for $d = 0, \cdots, 3$, as well as the normalized magnitude of the true acceleration. Clearly maneuvers are detected by the proposed smoothing algorithm ($d \geq 1$) more quickly compared to the forward-time IMM filtering algorithm. Except for a short period following model switching, the probability of one of the models in the model set is always quite close to one and it reflects the true motion status of the target, whereas the mode probability obtained via the forward-time IMM filtering

algorithm is sometimes somewhat uncertain even after the transient stage.

Overall it is seen that even a small lag can lead to a much better state estimation performance.

7 Conclusions

We investigated a suboptimal approach to the fixed-lag smoothing problem for Markovian switching systems. A fixed-lag smoothing algorithm was developed based on the concept of interacting multiple models [2]. The filtering and smoothing for the original system were integrated by introducing a state-augmented system whose current state vector consists of the current and delayed states (down to a fixed-lag d) of the original system. The fixed-lag mode probabilities given measurements up to the current time were approximated using a simple but effective method.

Simulation results for estimating the trajectory of a maneuvering target were presented to compare the performances of the proposed smoothing algorithm and the forward-time IMM filtering algorithm using an example from [7]. The performance of the fixed-lag IMM smoother was significantly better than that of the IMM filter. The performance of the proposed fixed-lag IMM smoothing algorithm quickly approaches that achieved by the fixed-interval smoothing algorithm of [7] with increasing lag; recall that we have used the example of [7] in Sec. 6. Compared to fixed-interval smoothers, the fixed-lag smoothers can be implemented in real-time with a small delay.

Overall we run n filters/smoothers in parallel where n is the number of models in the model set. The total computational load is roughly $(d+1)$ times that required by the forward-time IMM filtering for the original system. Given measurements up to time t_k, in addition to the smoothed state estimate at time t_{k-d} we also obtain the smoothed state estimates from t_{k-d+1} through t_{k-1} and the current state estimate at t_k without any extra

effort.

8 References

[1] J.K. Tugnait, "Detection and estimation for abruptly changing systems," *Automatica*, vol. 18, pp. 607-615, Sept. 1982.

[2] H.A.P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Automatic Control*, vol. AC-33, pp. 780-783, Aug. 1988.

[3] Y. Bar-Shalom and T.E. Fortmann, *Tracking and Data Association.* New York: Academic Press, 1988.

[4] Y. Bar-Shalom and X.R. Li, *Estimation and Tracking: Principles, Techniques and Software.* Norwood, MA: Artech House, 1993.

[5] V.J. Mathews and J.K. Tugnait, "Detection and estimation with fixed lag for abruptly changing systems," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-19, pp. 730-739, Sept. 1983.

[6] H.A.P. Blom and Y. Bar-Shalom, "Time-reversion of a hybrid state stochastic difference system with a jump-linear smoothing application," *IEEE Trans. Information Theory*, vol. IT-36, pp. 836-847, July 1990.

[7] R.E. Helmick, W.D. Blair and S.A. Hoffman, "Fixed-interval smoothing for Markovian switching systems," *IEEE Trans. Information Theory*, vol. IT-41, pp. 1845-1855, Nov. 1995.

[8] B.D.O. Anderson and J.B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[9] K.S. Narendra and J. Balakrishnan, "Adaptive control using multiple models," *IEEE Trans. Automatic Control*, vol. AC-42, pp. 171-187, Feb. 1997.

[10] X.R. Li and Y. Bar-Shalom, "Design of an interacting multiple model algorithm for air traffic control tracking," *IEEE Trans. Control Systems Technology*, vol. 1, pp. 186-194, Sept. 1993.

[11] X.R. Li, "Hybrid estimation techniques," pp. 1-76 in *Control & Dynamic Systems: Advances in Theory & Applications*, vol. 76, C.T. Leondes (Ed). New York: Academic, 1996.

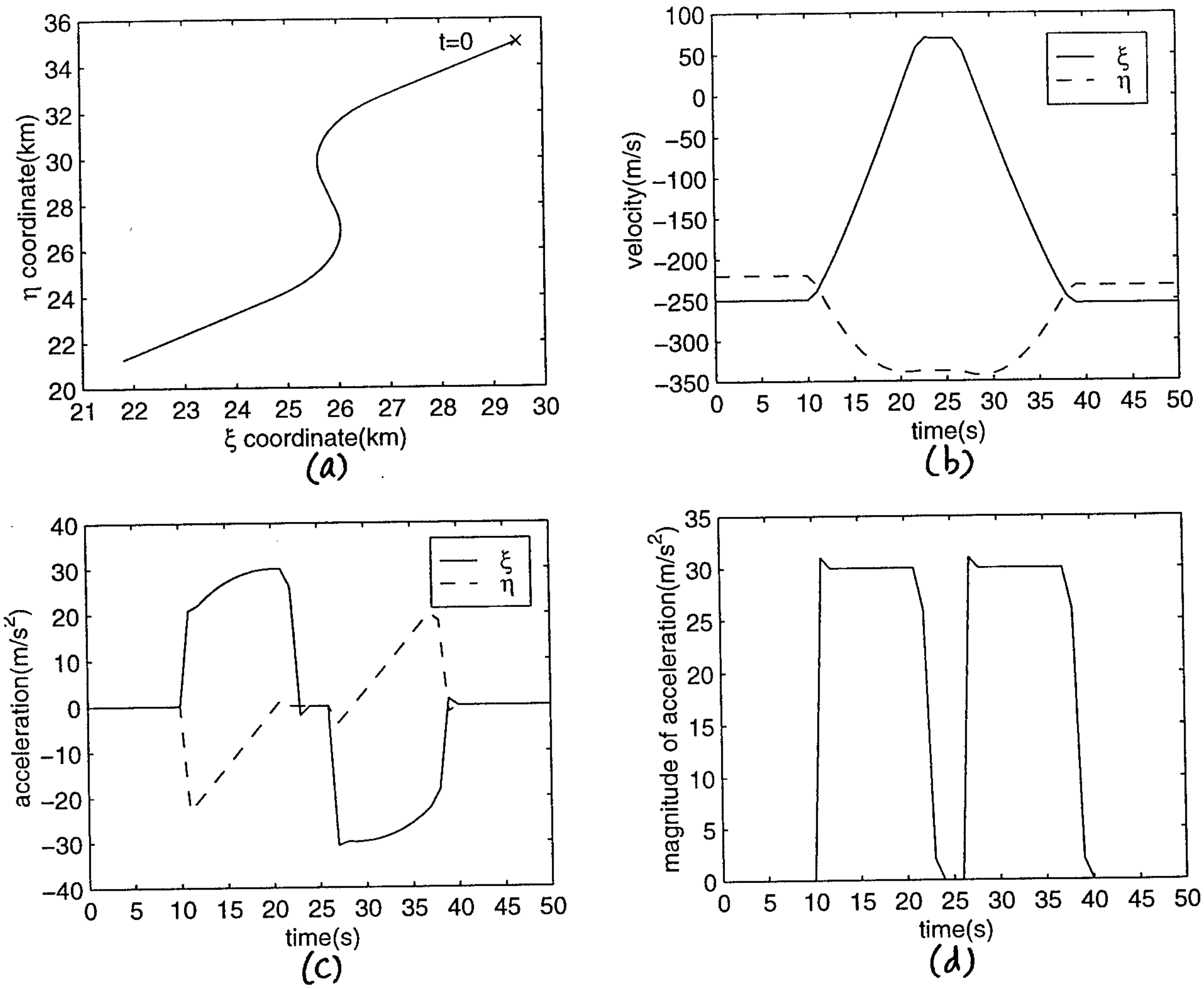


Fig. 1. The trajectory of the target: (a) Position in the $\xi\eta$-plane. (b) ξ and η velocities. (c) ξ and η accelerations. (d) The magnitude of the acceleration.

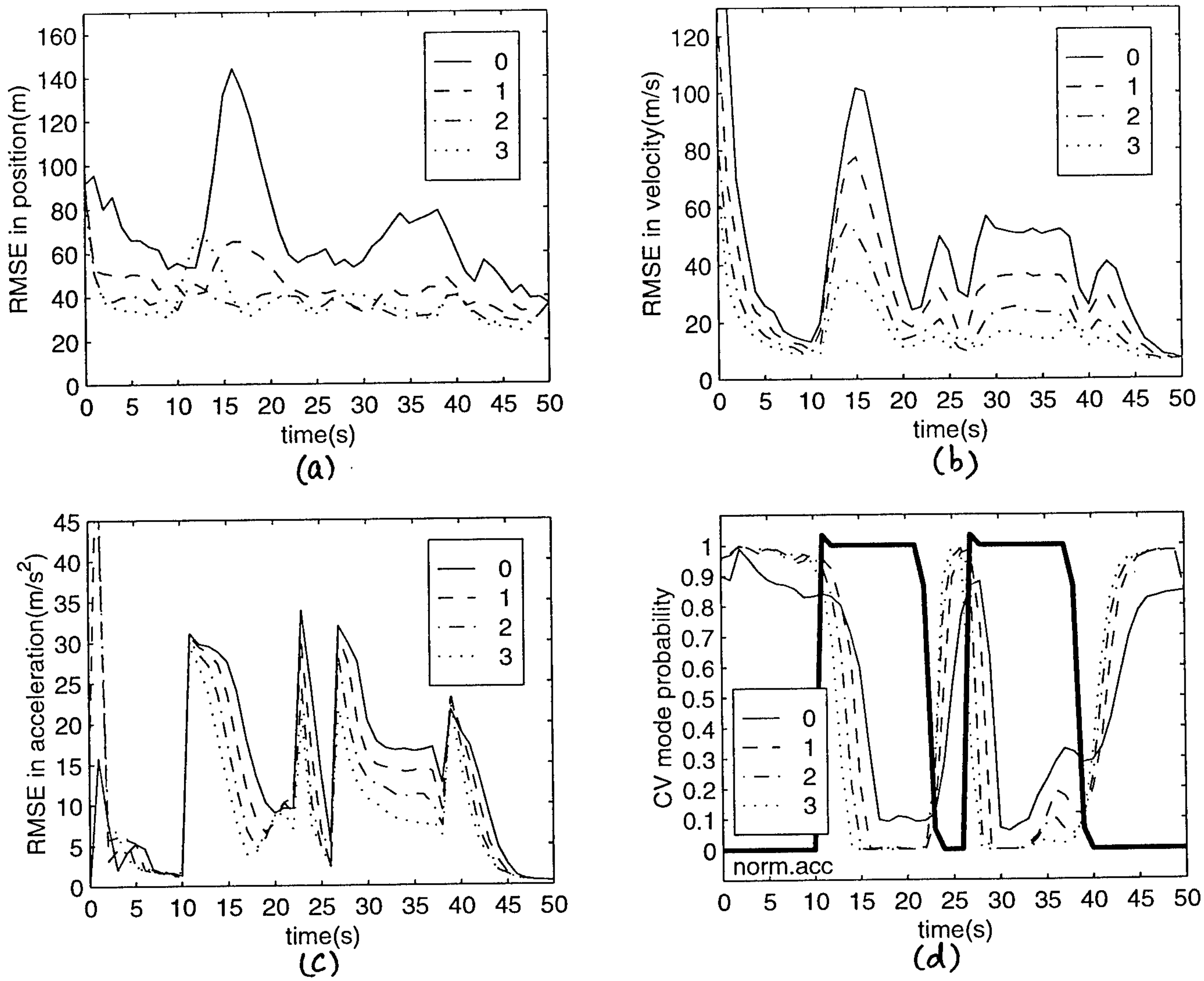


Fig. 2. Comparison of the filter and smoothers for various fixed-lags $d = 0, 1, 2, 3$ *: (a) RMSE in position. (b) RMSE in velocity. (c) RMSE in acceleration. (d) CV model probability.* Solid: lag d=0; dashed: lag d=1; dot-dashed: lag d=2; dotted: lag d=3.

[3]

Second-Order Statistics-Based Blind Equalization Of IIR Single-Input Multiple-Output Channels With Common Zeros[1]

Jitendra K. Tugnait & *Bin Huang*

Department of Electrical Engineering

Auburn University, Auburn, AL 36849, USA

Tel.: (334)844-1846 FAX: (334)844-1809

Email: tugnait@eng.auburn.edu

Abstract

The problem of blind equalization of SIMO (single-input multiple-output) communications channels is considered using only the second-order statistics of the data. Such models arise when a single receiver data is fractionally sampled (assuming that there is excess bandwidth), or when an antenna array is used with or without fractional sampling. We focus on direct design of finite-length MMSE (minimum mean-square error) blind equalizers. Unlike the past work on this problem, we allow infinite impulse response (IIR) channels. Our approaches also work when the "subchannel" transfer functions have common zeros so long as the common zeros are minimum-phase zeros. Illustrative simulation examples are provided.

[1]This work was supported by the National Science Foundation under Grant MIP-9312559 and by the the Office of Naval Research under Grant N00014-97-10822.

1 Introduction

Consider a discrete-time SIMO (single-input multiple-output) system with N outputs and one input. The i-th component of the output at time k is given by

$$y_i(k) = \mathcal{F}_i(z)w(k) + n_i(k), \quad i = 1, 2, \cdots, N, \tag{1-1}$$

$$\Longrightarrow \quad \mathrm{y}(k) = \mathcal{F}(z)w(k) + \mathrm{n}(k) = \mathrm{s}(k) + \mathrm{n}(k), \tag{1-2}$$

where $\mathrm{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathrm{s}(k)$ and $\mathrm{n}(k)$, and z is the $\mathcal{Z}-$transform variable as well as the backward-shift operator (i.e., $z^{-1}w(k) = w(k-1)$, etc.). The sequence $w(k)$ is the (single) input at sampling time k, $y_i(k)$ is the i-th noisy output, $s_i(k)$ is the i-th noise-free output, $n_i(k)$ is the additive measurement noise, and

$$F_i(z) := \sum_{l=0}^{\infty} f_i(l) z^{-l} \tag{1-3}$$

is the scalar transfer function with $w(k)$ as the input and $y_i(k)$ as the output; it represents the $i-$th subchannel. We allow all of the above variables to be complex-valued. The overall transfer function is denoted by the $N \times 1$ $\mathcal{F}(z)$ with its $i-$th element as $\mathcal{F}_i(z)$. We have

$$\mathcal{F}(z) = \sum_{i=0}^{\infty} \mathbf{F}_i z^{-i}. \tag{1-4}$$

Such models arise in several useful baseband-equivalent digital communications and other applications. A case of some interest is that of fractionally-spaced samples of a single baseband received signal leading to a SIMO model [1],[4],[8]. Alternatively, a similar model can be derived when we have a single signal impinging upon an antenna array with N elements [5]. A similar model arises if we have an antenna array coupled with fractional sampling at each array-element [5].

In these applications one of the objectives is to recover the inputs $w(k)$ given the noisy measurements but not given the knowledge of the system transfer function. Recently there has been much interest in solving this problem using only (or at least, to the maximum extent possible) the second-order statistics (SOS) of the data (see [1], [3]-[5], [8]-[14] and references therein). The solution is closely tied to existence of an FIR (finite impulse response) inverse to the system transfer function [1], [3]-[5], [8]-[14]. An overwhelming number of papers (see [4],[5],[9]-[12] and references

therein) have concentrated on a two-step procedure: first estimate the channel impulse response (IR) and then design an equalizer using the estimated channel. A fundamental restriction in these works is that the channel is FIR with no common zeros among the various subchannels. A few (see [1]and [13], e.g.) have proposed direct design of the equalizer bypassing channel estimation. Still they assume FIR channels with no common zeros.

In this paper we allow IIR (infinite impulse response) channels (which are finitely parametrized). We will also allow common zeros so long as they are minimum-phase (i.e., they lie inside the unit circle). Finally, in the presence of nonminimum-phase common zeros, our proposed approach equalizes the spectrally-equivalent minimum-phase counterpart of $\mathcal{F}(z)$; it does not "fall apart" unlike quite a few existing approaches. We should note that our proposed approach is inspired by [1]. Unlike [1] our approach applies to antenna arrays since we do not require that $f_1(0) \neq 0$ but $f_i(0) = 0$ for $i = 2, 3, \cdots, N$, as is required by [1]. This requirement of [1] is not restrictive for single-receiver causal systems with fractional sampling as one can always achieve this by "shifting," i.e. "re-grouping" of fractional samples per symbol. It does demand symbol synchronization so that fractional samples belonging to a given symbol are known thereby allowing for shifting or re-grouping to achieve the aforementioned requirement. In this paper we don't require such a synchronization; only the baud rate ought to be known.

Note that the prediction error methods of [8], [9] and [14] apply to the problem under consideration with some straightforward extensions/modifications (as we discuss in Sec. 3.3.3). Interestingly, [8], [9] and [14] derive their results under the assumption of FIR channels with no common zeros.

Three approaches are proposed in this paper for designing a blind MMSE (minimum mean-square error) linear equalizer of a specified length and delay. The approaches do not require the knowledge of the underlying system model orders or IR length. Algorithms I and II are inspired by [1] whereas Algorithm III is a straightforward extension of [9] and [14]. Algorithm II also exploits some results from [9] and [14] (see Remark 4 in Sec. 3.1).

The paper is organized as follows. Precise model assumptions and some background results used later in the paper are stated and developed in Sec. 2. IIR channels with no common subchannel zeros are considered in Sec. 3 where the three proposed algorithms of this paper are developed.

Under the assumptions of Sec. 3, finite length inverses and zero-forcing equalizers exist. In Sec. 4 we allow common subchannel zeros. Here ideally we need infinite length inverses and zero-forcing equalizers. Two computer simulation examples involving a 4-QAM signal are presented in Sec. 5 to illustrate and compare the performances of the proposed approaches.

2 Model Assumptions and Preliminaries

In this section we consider precise model assumptions and some background results used later in the paper. The material in Sec. 2.1 is useful in developing Algorithm I (see Sec. 3) whereas the material in Sec. 2.2 is useful in developing Algorithm III. Algorithm II exploits both Secs. 2.1 and 2.2. Lemma 2 is needed to estimate the noise variance. Lemma 1 is a straightforward extension of the results of [9] and [14].

2.1 FIR Inverses

Let $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ where $\mathcal{A}(z) = 1 + \sum_{i=1}^{n_a} a_i z^{-i}$ is 1×1 and $\mathcal{B}(z) = \sum_{i=0}^{n_b} \mathbf{B}_i z^{-i}$ is $N \times 1$. Assume the following:

- **(H1)** $N > 1$.
- **(H2)** $\text{Rank}\{\mathcal{B}(z)\} = 1 \ \forall z$ including $z = \infty$ but excluding $z = 0$, i.e., $\mathcal{B}(z)$ is irreducible [7, Sec. 6.3].
- **(H3)** $\mathcal{A}(z) \neq 0$ for $|z| \geq 1$.

Assumption **(H2)** is equivalent to stating that the various subchannels $F_i(z)$ have no common zeros. It has been shown in [6] (using some results from [2]) that under **(H1)**-**(H3)** there exists a finite degree left-inverse (not necessarily unique) of $\mathcal{F}(z)$:

$$\mathcal{G}(z)\mathcal{F}(z) = 1 \tag{2-1}$$

where $\mathcal{G}(z)$ is $1 \times N$ given by

$$\mathcal{G}(z) = \sum_{l=0}^{L_e} \mathbf{G}_l z^{-l} \quad \text{for any} \quad L_e \geq n_a + n_b - 1. \tag{2-2}$$

Remark 1: The left-inverse $\mathcal{G}(z)$ of $\mathcal{F}(z)$ consists of two parts: $\mathcal{G}(z) = \mathcal{G}_B(z)\mathcal{A}(z)$ where $\mathcal{G}_B(z)\mathcal{B}(z) = 1$ so that $\mathcal{G}(z)\mathcal{F}(z) = \mathcal{G}_B(z)\mathcal{A}(z)\mathcal{A}^{-1}(z)\mathcal{B}(z) = \mathcal{G}_B(z)\mathcal{B}(z) = 1$. Finite length left-inverses of FIR SIMO channels have been subject of intense research activities [4]-[6],[8]-[13]. Left-inverses to MIMO IIR/FIR channels have been considered in [6]. It appears that the results of [6] pertaining to MIMO models are the sharpest to date. Finally, it is important to stress that [4], [5] and [8]-[13] do not allow IIR channels, or subchannels having common zeros, in their problem formulation unlike this contribution.

2.2 Linear Innovations Representations

Assume further the following:

(H4) $\{w(k)\}$ is zero-mean, white. Take $E\{|w(k)|^2\} = 1$ by absorbing any non-identity correlation of $w(k)$ into $\mathcal{F}(z)$.

Lemma 1. Under (H1)-(H4), $\{\mathrm{s}(k)\}$ may be represented as

$$\mathrm{s}(k) = -\sum_{i=1}^{M} \mathbf{D}_i \mathrm{s}(k-i) + I_s(k) \tag{2-3}$$

where $M = n_a + n_b - 1$, $\mathbf{D}_i$'s are some $N \times N$ matrices such that $\det(\mathcal{D}(z)) \neq 0$ for $|z| \geq 1$, $\mathcal{D}(z) = I + \sum_{i=1}^{M} \mathbf{D}_i z^{-i}$ and $\{I_s(k)\}$ is a zero-mean white $N \times 1$ random sequence (linear innovations for $\{\mathrm{s}(k)\}$) with

$$E\{I_s(k)I_s^{\mathcal{H}}(k)\} = \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} \text{ and } \|\mathbf{F}_0\|^{-2}\mathbf{F}_0^{\mathcal{H}} I_s(k) = w(k). \quad \bullet \tag{2-4}$$

Proof: Consider the process

$$\mathrm{s}'(k) := \mathcal{A}(z)\mathrm{s}(k) = \mathcal{B}(z)w(k). \tag{2-5}$$

By [9] and [14], under **(H1)**, **(H2)** and **(H4)**, we have

$$\mathrm{s}'(k) = -\sum_{i=1}^{n_b-1} \mathbf{D}_i' \mathrm{s}'(k-i) + I_s'(k) \tag{2-6}$$

where $\mathbf{D}_i'$s are some $N \times N$ matrices such that $\det(\mathcal{D}'(z)) \neq 0$ for $|z| \geq 1$, $\mathcal{D}'(z) = I + \sum_{i=1}^{M} \mathbf{D}_i' z^{-i}$ and $\{I_s'(k)\}$ is a zero-mean white $N \times 1$ random sequence (linear innovations for $\{\mathrm{s}'(k)\}$) with

$$E\{I_s'(k)I_s'^{\mathcal{H}}(k)\} = \mathbf{B}_0\mathbf{B}_0^{\mathcal{H}} = \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} \text{ and } \|\mathbf{F}_0\|^{-2}\mathbf{F}_0^{\mathcal{H}} I_s'(k) = w(k). \tag{2-7}$$

Since $s(k) = \mathcal{A}^{-1}(z)s'(k)$, it follows from (2-6) that (2-3) holds true with $I_s(k) \equiv I'_s(k)$ such that $\mathcal{D}(z) = \mathcal{A}(z)\mathcal{D}'(z)$. This completes the proof. □

Lemma 2. Let $\mathcal{R}_{ssL_e}$ denote a $[N(L_e+1)] \times [N(L_e+1)]$ matrix with its ij-th block element as $\mathbf{R}_{ss}(j-i) = E\{s(k+j-i)s^{\mathcal{H}}(k)\}$. Then under (H1)-(H4), $\rho(\mathcal{R}_{ssL_e}) \le NL_e + 1$ for $L_e \ge n_a + n_b - 1$ where $\rho(A)$ denotes the rank of A. •

Proof: It follows from Lemma 1 and (2-3) that

$$\begin{bmatrix} I & \mathbf{D}_1 & \cdots & \mathbf{D}_{n_a+n_b-1} & 0 & \cdots & 0 \end{bmatrix} \mathcal{R}_{ssL_e} = \begin{bmatrix} \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} & 0 & \cdots & 0 \end{bmatrix}. \tag{2-8}$$

Clearly

$$\rho\left(\begin{bmatrix} I & \mathbf{D}_1 & \cdots & \mathbf{D}_{n_a+n_b-1} & 0 & \cdots & 0 \end{bmatrix}\right) = N \tag{2-9}$$

and

$$\rho\left(\begin{bmatrix} \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} & 0 & \cdots & 0 \end{bmatrix}\right) = 1. \tag{2-10}$$

Using (2-8)-(2-10) and Sylvester's inequality [7, p. 655], it follows that

$$\rho(\mathcal{R}_{ssL_e}) + N - N(L_e+1) \le 1 \tag{2-11}$$

which yields the desired result. □

3 Blind Equalization: No Common Zeros

In this section IIR channels with no common subchannel zeros are considered. For these channels finite length inverses and zero-forcing equalizers exist. The main objective of this paper is to design a blind MMSE linear equalizer of a specified length and delay. To this end, as will become clear in Sec. 3.2, we need to consider the design of a zero-forcing zero-delay linear equalizer of a specified length which is discussed in Sec. 3.1.

Assume that assumptions **(H1)**-**(H4)** hold true. In addition assume the following regarding the measurement noise:

(H5) $\{\mathbf{n}(k)\}$ is zero-mean with $E\{\mathbf{n}(k+\tau)\mathbf{n}^{\mathcal{H}}(k)\} = \sigma_n^2 I_{N\times N}$ where $I_{N\times N}$ is the $N \times N$ identity matrix.

3.1 Zero-Delay Zero-Forcing Equalizer

Using (2-1) and (2-2) and setting $\mathcal{F}(z) = \sum_{i=0}^{\infty} \mathbf{F}_i z^{-i}$, we have

$$\sum_{l=0}^{\infty} \mathbf{G}_{m-l}\mathbf{F}_l = \begin{cases} 1, & m = 0 \\ 0, & m = 1, 2, \cdots, \end{cases} \tag{3-1}$$

leading to

$$\begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \cdots & \mathbf{G}_{L_e} \end{bmatrix} \overline{\mathcal{S}} = \begin{bmatrix} 1 & 0 & \cdots & \cdots \end{bmatrix} \tag{3-2}$$

where $\overline{\mathcal{S}}$ is the $(N(L_e+1)) \times \infty$ matrix given by

$$\overline{\mathcal{S}} = \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3 & \cdots & \cdots & \cdots \\ 0 & \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & & & \\ 0 & 0 & \cdots & 0 & \mathbf{F}_0 & \mathbf{F}_1 & \cdots \end{bmatrix}. \tag{3-3}$$

Let $\overline{\mathcal{S}}^{\#}$ denote the pseudoinverse of $\overline{\mathcal{S}}$. By [15, Prop. 1], $\overline{\mathcal{S}}^{\#} = \overline{\mathcal{S}}^{\mathcal{H}}(\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}})^{\#}$. Then the minimum norm solution to the FIR equalizer is given by [15, Sec. 6.11]

$$\begin{aligned} \begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \cdots & \mathbf{G}_{L_e} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & \cdots & \cdots \end{bmatrix} \overline{\mathcal{S}}^{\#} \\ &= \begin{bmatrix} \mathbf{F}_0^{\mathcal{H}} & 0 & \cdots & 0 \end{bmatrix} (\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}})^{\#}. \end{aligned} \tag{3-4}$$

In a fashion similar to $\mathcal{R}_{ssL_e}$ in Lemma 2, let $\mathcal{R}_{yyL_e}$ denote a $[N(L_e+1)] \times [N(L_e+1)]$ matrix with its ij-th block element as $\mathbf{R}_{yy}(j-i) = E\{\mathbf{y}(k+j-i)\mathbf{y}^{\mathcal{H}}(k)\}$; define similarly $\mathcal{R}_{nnL_e}$ pertaining to the additive noise. Carry out an eigendecomposition of $\mathcal{R}_{yyL_e}$. Then the smallest $N-1$ eigenvalues of $\mathcal{R}_{yyL_e}$ equal σ_n^2 because under **(H1)**-**(H4)**, $\rho(\mathcal{R}_{ssL_e}) \leq NL_e + 1$ whereas $\rho(\mathcal{R}_{nnL_e}) = NL_e + N = \rho(\mathcal{R}_{yyL_e})$. Thus a consistent estimate $\widehat{\sigma}_n^2$ of σ_n^2 is obtained by taking it as the average of the smallest $N-1$ eigenvalues of $\widehat{\mathcal{R}}_{yyL_e}$, the data-based consistent estimate of $\mathcal{R}_{yyL_e}$.

Under **(H4)** and **(H5)**,

$$(\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}}) = \mathcal{R}_{ssL_e} = \mathcal{R}_{yyL_e} - \mathcal{R}_{nnL_e} = \mathcal{R}_{yyL_e} - \sigma_n^2 I. \tag{3-5}$$

Thus, $(\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}})$ can be estimated from noisy data. However, we don't know $\mathbf{F}_0$. To this end, we seek an $N \times N$ FIR filter $\mathcal{G}_a(z) := \sum_{i=0}^{L_e} \mathbf{G}_{ai} z^{-i}$ satisfying

$$\left[\begin{array}{cccc} \mathbf{G}_{a0} & \mathbf{G}_{a1} & \cdots & \mathbf{G}_{aL_e} \end{array} \right] = \left[\begin{array}{cccc} I_{N\times N} & 0 & \cdots & 0 \end{array} \right] \mathcal{R}^{\#}_{ssL_e}. \tag{3-6}$$

Comparing (3-4) and (3-6) it follows that

$$\left[\begin{array}{cccc} \mathbf{G}_0 & \mathbf{G}_1 & \cdots & \mathbf{G}_{L_e} \end{array} \right] = \mathbf{F}_0^{\mathcal{H}} \left[\begin{array}{cccc} \mathbf{G}_{a0} & \mathbf{G}_{a1} & \cdots & \mathbf{G}_{aL_e} \end{array} \right] \tag{3-7}$$

leading to

$$\sum_{i=0}^{L_e} \mathbf{G}_i z^{-i} =: \mathcal{G}(z) = \mathbf{F}_0^{\mathcal{H}} \mathcal{G}_a(z). \tag{3-8}$$

In practice, therefore, we apply $\mathcal{G}_a(z)$ to the data leading to

$$\mathrm{v}(k) := \mathcal{G}_a(z)\mathrm{y}(k) = \mathrm{v}_s(k) + \mathcal{G}_a(z)\mathrm{n}(k) \tag{3-9}$$

such that

$$\mathbf{F}_0^{\mathcal{H}} \mathrm{v}_s(k) = w(k) \tag{3-10}$$

where

$$\mathrm{v}_s(k) := \mathcal{G}_a(z)\left[\mathrm{y}(k) - \mathrm{n}(k)\right] = \mathcal{G}_a(z)\mathrm{s}(k). \tag{3-11}$$

In (3-10) $\{w(k)\}$ is a white scalar sequence (by assumption (**H4**)), however, $\{\mathrm{v}_s(k)\}$ is not necessarily a white vector sequence. Given the second-order statistics of $\{\mathrm{v}_s(k)\}$, how does one estimate $\mathbf{F}_0$ so that $\{w(k)\}$ satisfying (**H4**) is recovered? We need to have $R_{ww}(\tau) := E\{w(k+\tau)w^*(k)\} = 0$ for $|\tau| \neq 0$. By (3-9), $R_{ww}(\tau) = \mathbf{F}_0^{\mathcal{H}} R_{v_s v_s}(\tau)\mathbf{F}_0$. Define ($L > 0$ is some large integer)

$$\overline{R}_{v_s v_s} := \left[R^T_{v_s v_s}(-1) \;\; R^T_{v_s v_s}(-2) \;\; \cdots \;\; R^T_{v_s v_s}(-L) \right]^T \tag{3-12}$$

where $R_{v_s v_s}(\tau) := E\{v_s(k+\tau)v_s^{\mathcal{H}}(k)\}$.

Lemma 3. $\overline{R}_{v_s v_s}$ is rank deficient for any $L \geq 1$ such that $\overline{R}_{v_s v_s}\mathbf{F}_0 = 0$. •

Proof: We have

$$R_{wv_s}(\tau) = E\{w(k+\tau)v_s^{\mathcal{H}}(k)\} = 0 \quad \forall \tau \geq 1 \tag{3-13}$$

because $v_s(k)$ is obtained by causal filtering of $\mathbf{y}(k)$, hence of $w(k)$. Using (3-10) in (3-13) it then follows that there exists a $N \times 1$ $\mathbf{F}_0 \neq 0$ such that

$$\mathbf{F}_0^{\mathcal{H}} R_{v_s v_s}(\tau) = 0 \quad \forall \tau \geq 1. \tag{3-14}$$

Equivalently, we have from (3-14)

$$R_{v_s v_s}(-\tau)\mathbf{F}_0 = 0 \quad \forall \tau \geq 1. \tag{3-15}$$

The desired result is then immediate. □

Pick a $N \times 1$ column-vector $\mathbf{H}_0$ to equal the rightmost right singular vector in a singular-value decomposition (SVD) $\overline{R}_{v_s v_s} = U \Sigma V^{\mathcal{H}}$, i.e. the right singular vector corresponding to the smallest singular value. In other words, pick $\mathbf{H}_0$ to equal the last column of V. Then since ideally the smallest singular value of $\overline{R}_{v_s v_s}$ is zero, we have $\mathbf{H}_0^{\mathcal{H}} R_{v_s v_s}(-\tau)\mathbf{H}_0 = 0$ for $\tau = 1, 2, \cdots, L$. This, in turn, implies that

$$\left(\mathbf{H}_0^{\mathcal{H}} R_{v_s v_s}(-\tau)\mathbf{H}_0\right)^{\mathcal{H}} = \mathbf{H}_0^{\mathcal{H}} R_{v_s v_s}(\tau)\mathbf{H}_0 = 0 \text{ for } \tau = 1, 2, \cdots, L. \tag{3-16}$$

Since the overall system with $w(k)$ as input and $\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k)$ as output is ARMA$(n_a, n_b + L_e)$, it follows that $\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k)$ is zero-mean white if $L \geq n_b + L_e$, hence, a scaled version of $w(k)$. Therefore, we have $(\alpha \neq 0)$

$$\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k) =: w'(k) = \alpha w(k) \tag{3-17}$$

(because $\overline{R}_{v_s v_s}\mathbf{H}_0 = 0$). Thus, once $\mathbf{H}_0$ is found, one has the complete inverse filter to recover a scaled version of $w(k)$ via a zero-forcing filter.

Remark 2: In [1] $\mathbf{F}_0^{\mathcal{H}}$ in (3-4) has been replaced with an $N-$row vector $[1 \;\; 0 \;\; \cdots \;\; 0]$. This requirement of [1] is not restrictive for single-receiver causal systems with fractional sampling as one can always achieve this by "shifting," i.e. "re-grouping" of fractional samples per symbol: set $f_1(0) \neq 0$ but $f_i(0) = 0$ for $i = 2, 3, \cdots, N$. It does demand symbol synchronization so that fractional samples belonging to a given symbol are known thereby allowing for shifting or re-grouping to achieve the aforementioned requirement. In this paper we don't require such a

synchronization; only the baud rate ought to be known. The approach of [1] does not apply to antenna arrays whereas our approach does. □

Remark 3: If the noise is colored with known color except for a scalar scale factor, then we can follow prewhitening (as in [5]) and convert the problem to one that obeys (H5). □

Remark 4: $\mathbf{F}_0$ can also be estimated (up to a scale factor as unit norm $\mathbf{H}_0$) using the prediction error method of [9],[14] (even though [9] and [14] restrict their discussion to FIR models and real-valued data). Using (2-3) we obtain ($L_e \geq n_a + n_b - 1$)

$$\left[\mathbf{D}_1 \;\; \mathbf{D}_2 \;\; \cdots \;\; \mathbf{D}_{L_e} \right] \mathcal{R}_{ss(L_e-1)} = -\left[\mathbf{R}_{ss}(1) \;\; \mathbf{R}_{ss}(2) \;\; \cdots \;\; \mathbf{R}_{ss}(L_e) \right] \tag{3-18}$$

leading to the minimum norm solution

$$\left[\mathbf{D}_1 \;\; \mathbf{D}_2 \;\; \cdots \;\; \mathbf{D}_{L_e} \right] = -\left[\mathbf{R}_{ss}(1) \;\; \mathbf{R}_{ss}(2) \;\; \cdots \;\; \mathbf{R}_{ss}(L_e) \right] \mathcal{R}^{\#}_{ss(L_e-1)}. \tag{3-19}$$

Note that if $L_e > n_a + n_b - 1$, then $\mathbf{D}_i = 0$ for all $i > n_a + n_b - 1$ by virtue of Lemma 2. By (2-3) and (2-4) we also have

$$\mathbf{R}_{II}(0) := E\{I_s(k) I_s^{\mathcal{H}}(k)\} = \mathbf{F}_0 \mathbf{F}_0^{\mathcal{H}} = \mathbf{R}_{ss}(0) + \sum_{i=1}^{L_e} \mathbf{D}_i \mathbf{R}_{ss}(-i). \tag{3-20}$$

Clearly $\rho(\mathbf{R}_{ss}(0)) = 1$. Carry out an eigendecomposition of $\mathbf{R}_{II}(0)$. Pick $\mathbf{H}_0$ as the unit norm eigenvector corresponding to the largest eigenvalue (ideally the only nonzero eigenvalue) of $\mathbf{R}_{II}(0)$. □

Remark 5: It is worth noting that although $\mathbf{F}_0^{\mathcal{H}} \mathbf{v}_s(k) = w(k)$ (see (3-10)) and $\|\mathbf{F}_0\|^{-2} \mathbf{F}_0^{\mathcal{H}} I_s(k) = w(k)$ (see (2-4)), $\{I_s(k)\}$ is zero-mean white (linear innovations) whereas $\{\mathbf{v}_s(k)\}$ is in general colored. □

3.2 MMSE Equalizer with Delay d

We wish to design an MMSE (minimum mean-square error) linear equalizer of a specified length. It is not too hard to establish (using the orthogonality principle [16], for example) that the MMSE equalizer of length $L_e + 1$ to estimate $w(k-d)$ ($d \geq 0$) based upon $\mathbf{y}(n)$, $n = k, k-1, \cdots, k-L_e$, satisfies

$$\left[\overline{\mathbf{G}}_{d,0} \;\; \overline{\mathbf{G}}_{d,1} \;\; \cdots \;\; \overline{\mathbf{G}}_{d,L_e} \right] = \left[\mathbf{F}_d^{\mathcal{H}} \;\; \mathbf{F}_{d-1}^{\mathcal{H}} \;\; \cdots \;\; \mathbf{F}_0^{\mathcal{H}} \;\; 0 \;\; \cdots \;\; 0 \right] \mathcal{R}^{-1}_{yyL_e} \tag{3-21}$$

where $\mathcal{R}_{yyL_e}$ has its $ij-$th block-element given by $\mathbf{R}_{yy}(j-i)$. Clearly one can obtain a consistent estimate of $\mathcal{R}_{yyL_e}$ from the given data. It remains to estimate $\mathbf{F}_l$'s to complete the design. Here the discussion of Sec. 3.1 becomes relevant. There we found a $\mathbf{H}_0$ to satisfy (3-17). From (3-9) and (3-17) we have

$$\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k) = \sum_{i=0}^{L_e} \mathbf{H}_0^{\mathcal{H}}\mathbf{G}_{ai}\mathbf{s}(n-i). \tag{3-22}$$

Using (3-22) and taking expectations we have

$$E\{\mathbf{s}(n)\mathbf{v}_s^{\mathcal{H}}(n-\tau)\}\mathbf{H}_0 = \sum_{i=0}^{L_e} \mathbf{R}_{ss}(\tau+i)\mathbf{G}_{ai}^{\mathcal{H}}\mathbf{H}_0. \tag{3-23}$$

Using (1-2) and (3-17) we have

$$E\{\mathbf{s}(n)\mathbf{v}_s^{\mathcal{H}}(n-\tau)\}\mathbf{H}_0 = \alpha\mathbf{F}_\tau. \tag{3-24}$$

Hence, we have from (3-23) and (3-24)

$$\mathbf{F}_\tau^{\mathcal{H}} = \alpha^{-1}\mathbf{H}_0^{\mathcal{H}} \sum_{i=0}^{L_e} \mathbf{G}_{ai}\mathbf{R}_{ss}^{\mathcal{H}}(\tau+i). \tag{3-25}$$

Let $\mathcal{R}_{d,ssL_e}$ denote a $[N(L_e+1)] \times [N(L_e+1)]$ matrix with its ij-th block element as $E\{\mathbf{s}(k+d+j-i)\mathbf{s}^{\mathcal{H}}(k)\}$. Then (3-25) can be expressed as

$$\begin{bmatrix} \mathbf{F}_d^{\mathcal{H}} & \mathbf{F}_{d-1}^{\mathcal{H}} & \cdots & \mathbf{F}_0^{\mathcal{H}} & 0 & \cdots & 0 \end{bmatrix} = \alpha^{-1}\mathbf{H}_0^{\mathcal{H}} \begin{bmatrix} \mathbf{G}_{a0} & \mathbf{G}_{a1} & \cdots & \mathbf{G}_{aL_e} \end{bmatrix} \mathcal{R}_{d,ssL_e}^{\mathcal{H}}. \tag{3-26}$$

Finally, using (3-6) and (3-26) in (3-21) we obtain the desired solution

$$\begin{bmatrix} \overline{\mathbf{G}}_{d,0} & \overline{\mathbf{G}}_{d,1} & \cdots & \overline{\mathbf{G}}_{d,L_e} \end{bmatrix} = \alpha^{-1}\mathbf{H}_0^{\mathcal{H}} \begin{bmatrix} I_{N\times N} & 0 & \cdots & 0 \end{bmatrix} \mathcal{R}_{ssL_e}^{\#}\mathcal{R}_{d,ssL_e}^{\mathcal{H}}\mathcal{R}_{yyL_e}^{-1}. \tag{3-27}$$

The MMSE estimate $\widehat{w}(t-d)$ of $w(t-d)$ is then given by

$$\widehat{w}(t-d) = \sum_{i=0}^{L_e} \overline{\mathbf{G}}_{d,i}\mathbf{y}(t-i) \tag{3-28}$$

In practice, since α is unknown, one obtains a scaled version

$$\widetilde{w}(t-d) = \sum_{i=0}^{L_e} \alpha\overline{\mathbf{G}}_{d,i}\,\mathbf{y}(t-i) = \alpha\widehat{w}(t-d). \tag{3-29}$$

3.3 Algorithms: Practical Implementation

Given data $\mathbf{y}(k)$, $k = 1, 2, \cdots, T$. Pick the length $L_e + 1$ and delay d of the MMSE equalizer.

3.3.1 ALGORITHM I :

Here $\mathbf{F}_0$ is estimated as the unit norm $\mathbf{H}_0$ that lies in the null space of $\overline{R}_{v_s v_s}$.

I.1 Estimate the correlation function of the measurements at lag m as

$$\widehat{\mathbf{R}}_{yy}(m) = \frac{1}{T}\sum_{k=1}^{T} \mathbf{y}(k+m)\mathbf{y}^{\mathcal{H}}(k) \tag{3-30}$$

where we take $\mathbf{y}(k+m) = 0$ if $k+m < 1$ or $> T$. Define the $[N(L_e+1)] \times [N(L_e+1)]$ matrix $\widehat{\mathcal{R}}_{yyL_e}$ with its ij-th block element as $\widehat{\mathbf{R}}_{yy}(j-i)$. Carry out an eigendecomposition of $\widehat{\mathcal{R}}_{yyL_e}$. Let λ_i $(i = NL_e+2, \cdots, NL_e+N)$ denote the smallest $N-1$ eigenvalues of $\widehat{\mathcal{R}}_{yyL_e}$. Estimate the noise variance σ_n^2 as

$$\widehat{\sigma}_n^2 = \frac{1}{N-1}\sum_{i=NL_e+2}^{NL_e+N} \lambda_i. \tag{3-31}$$

The signal correlation function at lag m is then estimated as

$$\widehat{\mathbf{R}}_{ss}(m) = \widehat{\mathbf{R}}_{yy}(m) - \widehat{\sigma}_n^2 I_{N\times N}\delta(m) \tag{3-32}$$

where $\delta(m)$ is the Kronecker delta function. Define the $[N(L_e+1)] \times [N(L_e+1)]$ signal correlation matrix estimate $\widehat{\mathcal{R}}_{ssL_e}$ with its ij-th block element as $\widehat{\mathbf{R}}_{ss}(j-i)$.

I.2 Now we implement (3-6). First we need to calculate $\mathcal{R}^{\#}_{ssL_e}$. Carry out a singular value decomposition of $\widehat{\mathcal{R}}_{ssL_e}$ leading to $\widehat{\mathcal{R}}_{ssL_e} = \mathbf{U}\Sigma\mathbf{V}^{\mathcal{H}}$ where $\Sigma = diag\{s_i, i = 1, 2, \cdots, NL_e + N\}$. The rank n_1 of $\widehat{\mathcal{R}}_{ssL_e}$ is determined as the smallest n for which

$$q_n := \sqrt{\frac{\sum_{i=n+1}^{NL_e+N} s_i}{\sum_{i=1}^{NL_e+N} s_i}} \le \epsilon_1 \tag{3-33}$$

where $\epsilon_1 > 0$ is a small number. [For simulations presented in Sec. 5 we took $\epsilon_1 = 0.001$]. The desired pseudoinverse is then calculated as

$$\widehat{\mathcal{R}}^{(1)\#}_{ssL_e} = \mathbf{V}_1\Sigma_1^{-1}\mathbf{U}_1^{\mathcal{H}} \tag{3-34}$$

where $\Sigma_1 = diag\{s_i, i = 1, 2, \cdots, n_1\}$ and $\mathbf{U}_1$ and $\mathbf{V}_1$ are comprised of the left and the right (respectively) singular vectors corresponding to the singular values retained in Σ_1. Using (3-34) calculate

$$\left[\begin{array}{cccc} \widehat{\mathbf{G}}_{a0} & \widehat{\mathbf{G}}_{a1} & \cdots & \widehat{\mathbf{G}}_{aL_e} \end{array} \right] = \left[\begin{array}{cccc} I_{N\times N} & 0 & \cdots & 0 \end{array} \right] \widehat{\mathcal{R}}_{ssL_e}^{(1)\#}. \tag{3-35}$$

I.3 Using (3-11) estimate $\mathbf{R}_{v_s v_s}(m)$ as

$$\widehat{\mathbf{R}}_{v_s v_s}(m) = \sum_{l_1=0}^{L_e} \sum_{l_2=0}^{L_e} \widehat{\mathbf{G}}_{al_1} \widehat{\mathbf{R}}_{ss}(m + l_2 - l_1) \widehat{\mathbf{G}}_{al_2}^{\mathcal{H}} \tag{3-36}$$

where $\widehat{\mathbf{R}}_{ss}(m)$ has been discussed in (3-30)-(3-32). Define the $(LN) \times N$ matrix

$$\widehat{\widehat{\mathcal{R}}}_{v_s v_s} := \left[\widehat{R}_{v_s v_s}^T(1) \;\; \widehat{R}_{v_s v_s}^T(2) \;\; \cdots \;\; \widehat{R}_{v_s v_s}^T(L) \right]^T. \tag{3-37}$$

Carry out an SVD of $\widehat{\widehat{\mathcal{R}}}_{v_s v_s}$ and pick

$$\widehat{\mathbf{H}}_0 = \text{‘rightmost’ right singular vector of } \widehat{\widehat{\mathcal{R}}}_{v_s v_s}. \tag{3-38}$$

I.4 Define the $[N(L_e+1)] \times [N(L_e+1)]$ matrix $\widehat{\mathcal{R}}_{d,ssL_e}$ with its ij-th block element as $\widehat{\mathbf{R}}_{ss}(j-i)$. The MMSE equalizer with delay d is calculated as

$$\left[\begin{array}{cccc} \widehat{\widehat{\mathbf{G}}}_{d,0} & \widehat{\widehat{\mathbf{G}}}_{d,1} & \cdots & \widehat{\widehat{\mathbf{G}}}_{d,L_e} \end{array} \right] = \widehat{\mathbf{H}}_0^{\mathcal{H}} \left[\begin{array}{cccc} I_{N\times N} & 0 & \cdots & 0 \end{array} \right] \widehat{\mathcal{R}}_{ssL_e}^{(2)\#} \widehat{\mathcal{R}}_{d,ssL_e}^{\mathcal{H}} \widehat{\mathcal{R}}_{yyL_e}^{-1}. \tag{3-39}$$

In (3-39) $\widehat{\mathcal{R}}_{ssL_e}^{(2)\#}$ is the pseudoinverse of $\widehat{\mathcal{R}}_{ssL_e}$ calculated as in (3-34) except that a larger error threshold ϵ_2 is used in (3-33) instead of ϵ_1. The rank n_2 of $\widehat{\mathcal{R}}_{ssL_e}$ is determined as the smallest n for which $q_n \leq \epsilon_2$ in (3-33) instead of $q_n \leq \epsilon_1$. [For simulations presented in Sec. 5 we took $\epsilon_2 = 0.01$].

Remark 6. In (3-39) calculation of $\widehat{\mathcal{R}}_{ssL_e}^{(2)\#}$ is related to computation of some of the leading coefficients of the channel impulse response whereas in (3-35) calculation of $\widehat{\mathcal{R}}_{ssL_e}^{(1)\#}$ is related to the calculation of the null space of $\widehat{\widehat{\mathcal{R}}}_{v_s v_s}$. Heuristically, a higher value of ϵ in (3-33) leads to higher "intersymbol interference" but lower "noise enhancement" in a zero-forcing equalizer design, and vice-versa. In estimating $\mathbf{H}_0$ via (3-38) suppression of intersymbol interference is more important

in order to 'better define' the null space of $\widehat{\widehat{\mathcal{R}}}_{v_s v_s}$. In contrast, in (3-39) (also recall (3-6) and (3-26)) one requires a compromise between intersymbol interference and noise enhancement while estimating some of the channel impulse response coefficients. □

3.3.2 ALGORITHM II :

Here $\mathbf{F}_0$ is estimated as in Remark 4.

II.1 Repeat step I.1 of Algorithm I.

II.2 Calculate the pseudoinverse $\widehat{\mathcal{R}}^{(2)\#}_{ss(L_e-1)}$ of $\widehat{\mathcal{R}}_{ss(L_e-1)}$ as in step I.4 of Algorithm I. Calculate

$$\left[\begin{array}{cccc} \widehat{\mathbf{D}}_1 & \widehat{\mathbf{D}}_2 & \cdots & \widehat{\mathbf{D}}_{L_e} \end{array}\right] = -\left[\begin{array}{cccc} \widehat{\mathbf{R}}_{ss}(1) & \widehat{\mathbf{R}}_{ss}(2) & \cdots & \widehat{\mathbf{R}}_{ss}(L_e) \end{array}\right] \widehat{\mathcal{R}}^{(2)\#}_{ss(L_e-1)}. \tag{3-40}$$

Further calculate

$$\widehat{\mathbf{R}}_{II}(0) = \widehat{\mathbf{R}}_{ss}(0) + \sum_{i=1}^{L_e} \widehat{\mathbf{D}}_i \widehat{\mathbf{R}}_{ss}(-i). \tag{3-41}$$

Set $\widehat{\mathbf{H}}_0$ equal to the unit norm eigenvector corresponding to the largest eigenvalue of $\widehat{\mathbf{R}}_{II}(0)$.

II.3 Repeat step I.4 of Algorithm I with $\widehat{\mathbf{H}}_0$ obtained from step II.2.

3.3.3 ALGORITHM III :

Here we will use (3-21) with $\mathbf{F}_i$ $(i = 0, 1, \cdots, d)$ estimated using the basic approach of [9] and [14]. Although [9] and [14] derive all their results under the assumption of FIR channels with no common zeros, their results extend (with straightforward modifications) to models that satisfy **(H1)**-**(H5)** by virtue of Lemma 1. By (2-4), we have

$$w(k) = \|\mathbf{F}_0\|^{-2}\mathbf{F}_0^{\mathcal{H}} I_s(k) = \|\mathbf{F}_0\|^{-1}\mathbf{H}_0^{\mathcal{H}} I_s(k). \tag{3-42}$$

By (1-2) and (1-4) it follows that

$$s(k) = \sum_{i=0}^{\infty} \mathbf{F}_i w(k-i). \tag{3-43}$$

From (3-43) and (H4), we have the relations

$$E\{w(k-l)\mathbf{s}^{\mathcal{H}}(k)\} \;=\; \mathbf{F}_l^{\mathcal{H}} \quad \text{for} \quad l \geq 0. \tag{3-44}$$

From (2-3) and (3-42), we have the relations

$$E\{w(k-l)\mathbf{s}^{\mathcal{H}}(k)\} \;=\; \|\mathbf{F}_0\|^{-1}\mathbf{H}_0^{\mathcal{H}}\left[\mathbf{R}_{ss}(-l) + \sum_{i=1}^{L_e}\mathbf{D}_i\mathbf{R}_{ss}(-l-i)\right]. \tag{3-45}$$

From (3-44) and (3-45) it follows that

$$\mathbf{F}_l^{\mathcal{H}} \;=\; \|\mathbf{F}_0\|^{-1}\mathbf{H}_0^{\mathcal{H}}\left[\mathbf{R}_{ss}^{\mathcal{H}}(l) + \sum_{i=1}^{L_e}\mathbf{D}_i\mathbf{R}_{ss}^{\mathcal{H}}(l+i)\right]. \tag{3-46}$$

Based upon the above discussion, [9] and [14], we have the following algorithm:

III.1 Repeat step I.1 of Algorithm I.

III.2 Repeat step II.2 of Algorithm II.

III.3 Estimate $\mathbf{F}_l^{\mathcal{H}}$ up to a scale factor as

$$\widehat{\mathbf{F}}_l^{\mathcal{H}} \;=\; \widehat{\mathbf{H}}_0^{\mathcal{H}}\left[\widehat{\mathbf{R}}_{ss}^{\mathcal{H}}(l) + \sum_{i=1}^{L_e}\widehat{\mathbf{D}}_i\widehat{\mathbf{R}}_{ss}^{\mathcal{H}}(l+i)\right], \quad l = 0, 1, \cdots, d. \tag{3-47}$$

III.4 The MMSE equalizer of length $L_e + 1$ and with delay d is calculated (up to a scale factor) as

$$\left[\begin{array}{cccc} \widehat{\widehat{\mathbf{G}}}_{d,0} & \widehat{\widehat{\mathbf{G}}}_{d,1} & \cdots & \widehat{\widehat{\mathbf{G}}}_{d,L_e} \end{array}\right] = \left[\begin{array}{cccccccc} \widehat{\mathbf{F}}_d^{\mathcal{H}} & \widehat{\mathbf{F}}_{d-1}^{\mathcal{H}} & \cdots & \widehat{\mathbf{F}}_0^{\mathcal{H}} & 0 & \cdots & 0 \end{array}\right]\widehat{\mathcal{R}}_{yyL_e}^{-1}. \tag{3-48}$$

4 Blind Equalization: Common Zeros

Now we allow common subchannel zeros. In this case since ideally we need infinite length inverses and zero-forcing equalizers, the presented results hold true only approximately for finite length equalizers. Assume that **(H1)**-**(H5)** hold true.

4.1 Minimum-Phase Zeros

Here the SIMO transfer function is

$$\mathcal{F}(z) \;=\; \frac{\mathcal{B}_c(z)}{\mathcal{A}(z)}\mathcal{B}(z) \tag{4-1}$$

where $\mathcal{B}(z)$ satisfies (H2) and $\mathcal{B}_c(z)$ is a finite-degree scalar polynomial that collects all the common zeros of the subchannels. Assume that

(**H6**) Given model (4-1), $\mathcal{B}_c(z) \neq 0$ for $|z| \geq 1$.

Then while $\mathcal{A}^{-1}(z)\mathcal{B}(z)$ has a finite inverse, $\mathcal{B}_c^{-1}(z)$ is IIR though causal under (H6). Then (3-2) holds true approximately for "large" L_e, the approximation getting better with increasing L_e. Similarly Lemma 1 holds true approximately for "large" M and Lemma 2 also holds true approximately for $L_e \geq M$. It is then readily seen that the developments of Secs. 3.1, 3.2 and 3.3 apply to the current case also.

4.2 Arbitrary Zeros

In this case (4-1) is true but $\mathcal{B}_c(z)$ does not necessarily satisfy (**H6**). We may rewrite (4-1) as

$$\mathcal{F}(z) = \overline{\mathcal{F}}(z)\mathcal{F}_{AP}(z) \tag{4-2}$$

where $\mathcal{F}_{AP}(z)$ is an allpass (rational) function such that

$$\mathcal{B}_c(z)\mathcal{B}_c(z^{-1}) = \mathcal{F}_{AP}(z)\overline{\mathcal{B}}_{MP}(z) \tag{4-3}$$

and $\overline{\mathcal{B}}_{MP}(z)$ is minimum-phase. Thus (within a scale factor) we have

$$\overline{\mathcal{F}}(z) = \frac{\overline{\mathcal{B}}_{MP}(z)}{\mathcal{A}(z)}\mathcal{B}(z). \tag{4-4}$$

We may rewrite (1-2) as

$$\mathbf{y}(k) = \overline{\mathcal{F}}(z)w'(k) + \mathbf{n}(k) \tag{4-5}$$

where

$$w'(k) := \mathcal{F}_{AP}(z)w(k). \tag{4-6}$$

Clearly $w'(k)$ satisfies (**H4**). Hence, (4-4)-(4-6) satisfy the requirements of Sec. 4.1. Therefore, one can "approximately" recover $w'(k)$ from the given data by applying the algorithms of Sec. 3.3.

In order to recover $w(k)$ form $w'(k)$, one needs to exploit the higher-order statistics of $\{w'(k)\}$; see [2],[3] and references therein.

5 Simulation Examples

Here we consider two simulation examples to illustrate the proposed approaches. Both of the examples are modified versions of the example from [5]. Example 1 consists of an ARMA model whose MA part is taken from [5]. Example 2 consists of an MA (FIR) model where we augment the FIR channel of [5] with a zero at 0.5 where this zero is common to all of the four subchannels.

For computing $\widehat{\mathcal{R}}_{ssL_e}^{(1)\#}$ in (3-34) via SVD, we picked $\epsilon_1 = 0.001$ in (3-33). For computing $\widehat{\mathcal{R}}_{ssL_e}^{(2)\#}$ in (3-39), or $\widehat{\mathcal{R}}_{ss(L_e-1)}^{(2)\#}$ in (3-40), via SVD, we picked $\epsilon_2 = 0.01$ in (3-33). Moreover, $\widehat{\mathcal{R}}_{yyL_e}^{-1}$ in (3-39) and (3-48) was also computed using SVD where all singular values smaller than 0.001$\times$(largest singular value) were neglected. Thus, calculation of $\widehat{\mathcal{R}}_{yyL_e}^{-1}$ was regularized. The measurement SNR is defined as

$$\text{SNR} \;=\; \frac{\sum_{i=1}^{N} E\{|s_i(k)|^2\}}{\sum_{i=1}^{N} E\{|n_i(k)|^2\}}.$$

The normalized MSE (i.e., MSE divided by $E\{|w(k)|^2\}$) and the probability of detection error (P_e) after equalization were taken as the two performance measures after averaging over 100 Monte Carlo runs. The equalized data were rotated and scaled before calculating the two performance measures. After designing the equalizers based on the given data record, the designed equalizer was applied to an independent record of length 3000 symbols in order to calculate normalized MSE and P_e. Therefore, the estimated P_e is not reliable below approximately 10^{-4}, hence, these values are not shown in Figs. 2 and 4.

5.1 Example 1.

We have $N = 4$ in (1-2) with $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ where

$$\mathcal{A}(z) = (1 - 0.5z^{-1})I_{3\times 3} \tag{5-1}$$

and $\mathcal{B}(z)$ is 4×1 with its i−th element given by

$$\mathcal{B}_1(z) = (-0.049 + j0.359) + (0.482 - j0.569)z^{-1}$$

$$+(-0.556 + j0.587)z^{-2} + (1.0 + j0.0)z^{-3} + (-0.171 + j0.061)z^{-4}$$

$$\mathcal{B}_2(z) = (0.443 - j0.0364) + (1.0 + j0.0)z^{-1}$$

$$+(0.921 - j0.194)z^{-2} + (0.189 - j0.208)z^{-3} + (-0.087 - j0.054)z^{-4}$$

$$\mathcal{B}_3(z) = (-0.211 - j0.322) + (-0.199 + j0.918)z^{-1}$$

$$+(1.0 + j0.0)z^{-2} + (-0.284 - j0.524)z^{-3} + (0.136 - j0.190)z^{-4}$$

$$\mathcal{B}_4(z) = (0.417 + j0.030) + (1.0 + j0.0)z^{-1}$$

$$+(0.873 + j0.145)z^{-2} + (0.285 + j0.309)z^{-3} + (-0.049 + j0.161)z^{-4}. \tag{5-2}$$

The MA part $\mathcal{B}(z)$ is the same as the FIR channel of [5]. The scalar input $w(k)$ is 4-QAM (as in [5]).

Transfer function $\mathcal{B}(z)$ satisfies **(H2)** [5], therefore, there exists a finite left inverse of length $L_e = 4$ (cf. Sec. 2.1). An MMSE equalizer of length $L_e = 12$ (13 taps per subchannel, totaling 52 taps: substantial overfitting!!) was designed with a delay d =3 (arbitrarily selected just for illustration). The Algorithms I–III were applied for various record lengths. The equalized output was scaled to match the true $\{w(k)\}$ before computing the mean-square error (MSE) in the equalized output. Fig. 1 shows the normalized MSE and Fig. 2 shows the probability of error P_e, both averaged over 100 Monte Carlo runs. It is seen that the proposed design approaches can handle IIR channels with little difficulty. Algorithm II (newly proposed) performs the best with Algorithm III (based upon [9] and [14]) being almost as good. The performance of Algorithm I improves with increase in record length and it approaches that of the other two algorithms for $T = 1000$ symbols.

5.2 Example 2.

Again we have $N = 4$ in (1-2) but with $\mathcal{F}(z) = \mathcal{B}_c(z)\mathcal{B}(z)$ where $\mathcal{B}(z)$ is as in Example 1 and $\mathcal{B}_c(z)$ is a scalar polynomial given by

$$\mathcal{B}_c(z) \ = \ 1 - 0.5z^{-1}. \tag{5-3}$$

Thus all four subchannels have a common zero at 0.5. The input $w(k)$ is 4-QAM as in Example 1. Note that in this example a finite left inverse does not exist.

As in Example 1, an MMSE equalizer of length $L_e = 12$ was designed with a delay d =3. Fig. 3 shows the normalized MSE and Fig. 3 shows the probability of error P_e, both averaged over 100 Monte Carlo runs. It is seen that the proposed design approaches can handle subchannels with common minimum-phase zeros with little difficulty. As in Example 1, Algorithm II performs the best.

6 Conclusions

Direct blind MMSE equalization of SIMO channels using only the second-order statistics of the data was considered. Such channels arise when antenna arrays are used or when signals with excess bandwidth are fractionally sampled or when both these scenarios are applicable. Unlike the past work on this problem [4],[5],[8]-[14], the proposed solutions are applicable to IIR channels and to SIMO systems having common zeros among the various subchannels so long as the common zeros are minimum-phase. In case of nonminimum-phase zeros, we recover an allpass filtered version of the original input.

Three approaches were proposed. Algorithms I and II are inspired by [1] whereas Algorithm III is a straightforward extension of [9] and [14]. Algorithm II also exploits some results from [9] and [14]. Two illustrative simulation examples, one consisting of an IIR channel and the other consisting of an FIR channel with a common zero, were presented using a 4-QAM information sequence. The proposed approaches work well. Algorithm II works the best (evaluated in terms of mean-square error and probability of detection error after equalization) with Algorithm III being a close second.

Future work includes performance analysis, adaptive implementation and extension to MIMO scenarios involving more than one information signals.

7 References

[1] G.B. Giannakis and S.D. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: Adaptive and optimal solutions," in *Proc. 1995 IEEE ICASSP*, pp. 1972-1975, Detroit,

MI, May 9-12, 1995. [Also, to appear in *IEEE Trans. Signal Processing*, 1997.]

[2] J.K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 268-271, Jan. 1997.

[3] Special Issue, *IEEE Transactions on Signal Processing*, vol. SP-45, Jan. 1997.

[4] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channels," *IEEE Trans. Information Theory*, vol. IT-40, pp. 340-349, March 1994.

[5] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 516-525, Feb. 1995.

[6] J.K. Tugnait, "FIR inverses to MIMO rational transfer functions with application to blind equalization," in *Proc. 30th Annual Asilomar Conf. Signals Systems Computers*, pp. 295-299, Pacific Grove, CA, Nov. 1996.

[7] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[8] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction," in *Proc. 1994 IEEE ICASSP*, pp. IV:585-588, Adelaide, Australia, May 1994.

[9] K. Abed-Meraim et al. "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. 1995 IEEE ICASSP*, pp. 1968-1971, Detroit, MI, May 9-12, 1995.

[10] Z. Ding, "A blind channel identification algorithm based on matrix outer-product," in *Proc. 1996 IEEE ICC*, pp. 852-856, Dallas, TX, June 23-26, 1996.

[11] K. Abed-Meraim et al. "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 42-55, Jan. 1997.

[12] D. Gesbert and P. Duhamel, "Robust blind joint data/channel estimation based on bilinear optimization," in *Proc. Eighth IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 168-171, Corfu, Greece, June 24-26, 1996.

[13] D. Gesbert, P. Duhamel and S. Mayrargue, "Blind multichannel adaptive MMSE equalization with controlled delay," in *Proc. Eighth IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 172-175, Corfu, Greece, June 24-26, 1996.

[14] K. Abed-Meraim et al. "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 694-705, March 1997.

[15] D.G. Luenberger, *Optimization by Vector Space Methods.* New York: John Wiley, 1969.

[16] H.V. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag: New York, 1988.

FIGURE CAPTIONS

Fig. 1. Example 1: Normalized MSE after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

Fig. 2. Example 1: Probability of error after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

Fig. 3. Example 2: Normalized MSE after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

Fig. 4. Example 2: Probability of error after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

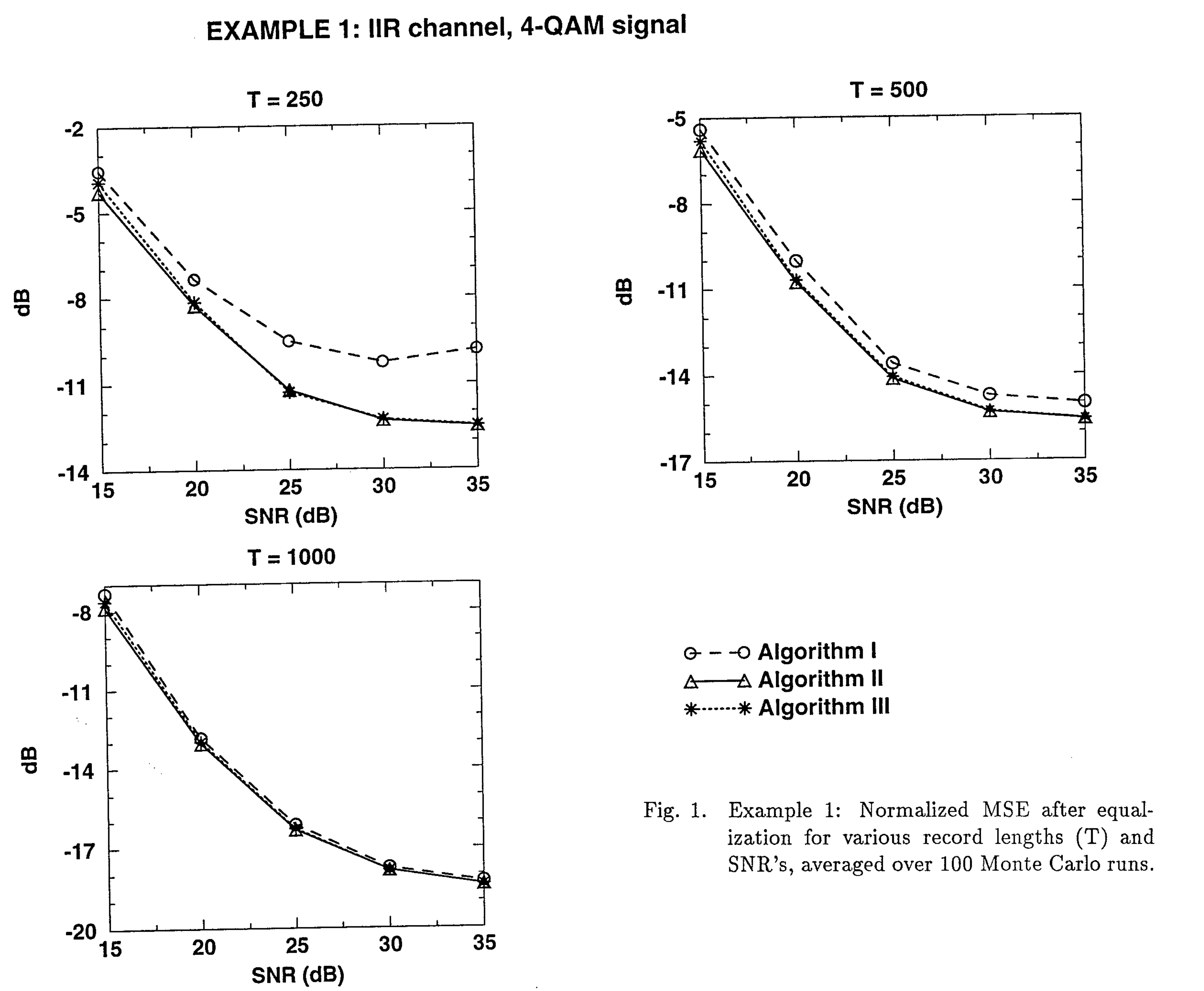


Fig. 1. Example 1: Normalized MSE after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

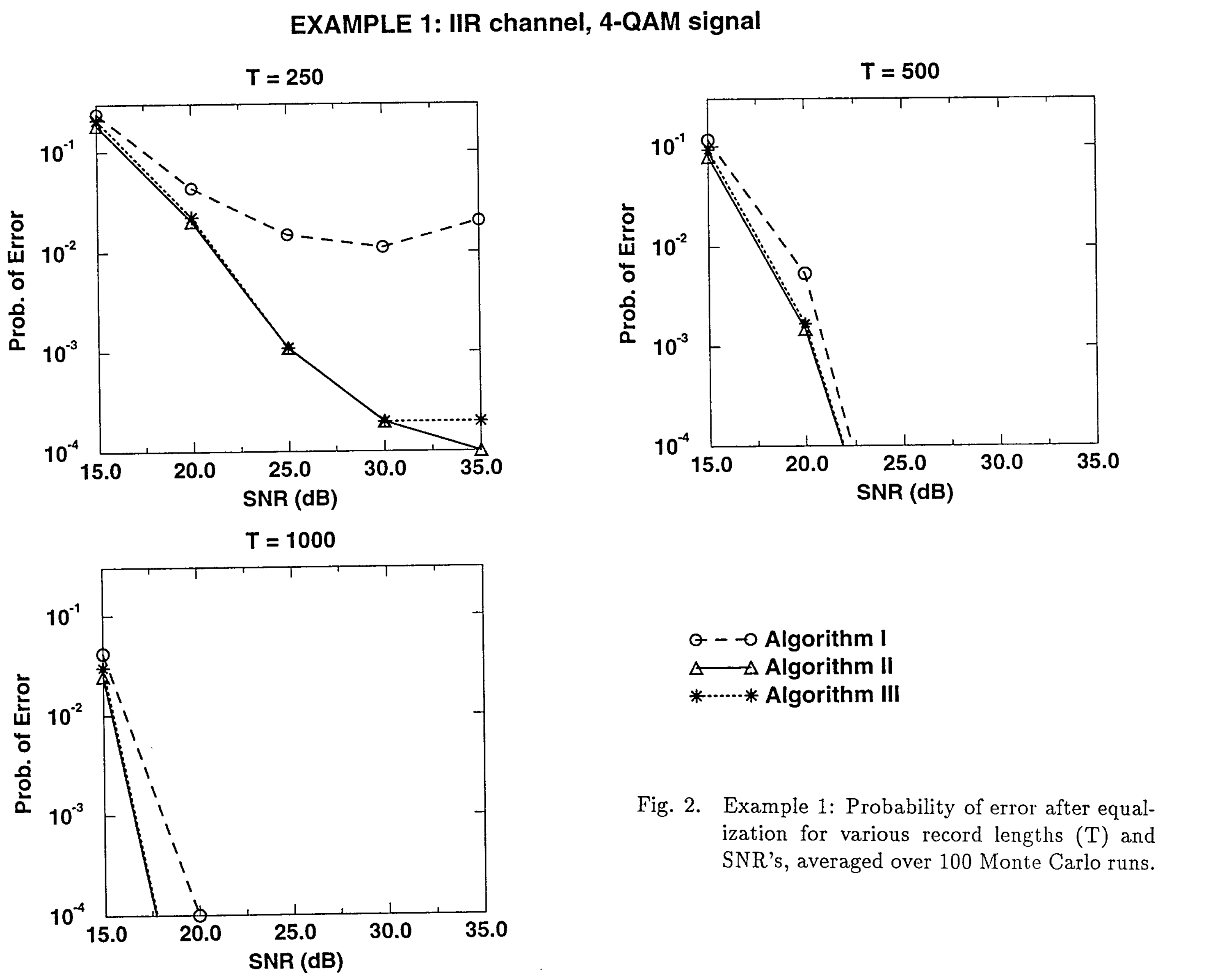


Fig. 2. Example 1: Probability of error after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

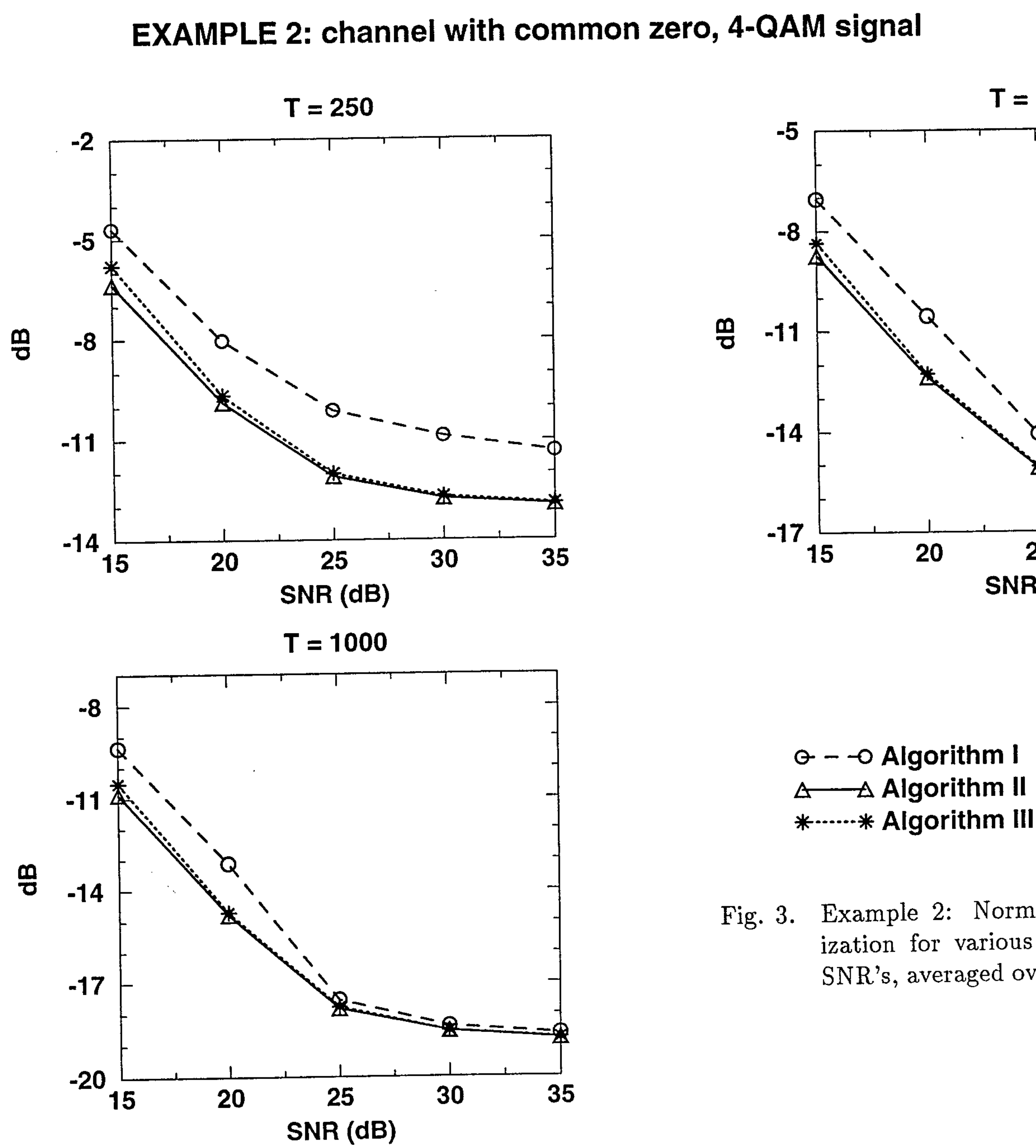


Fig. 3. Example 2: Normalized MSE after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

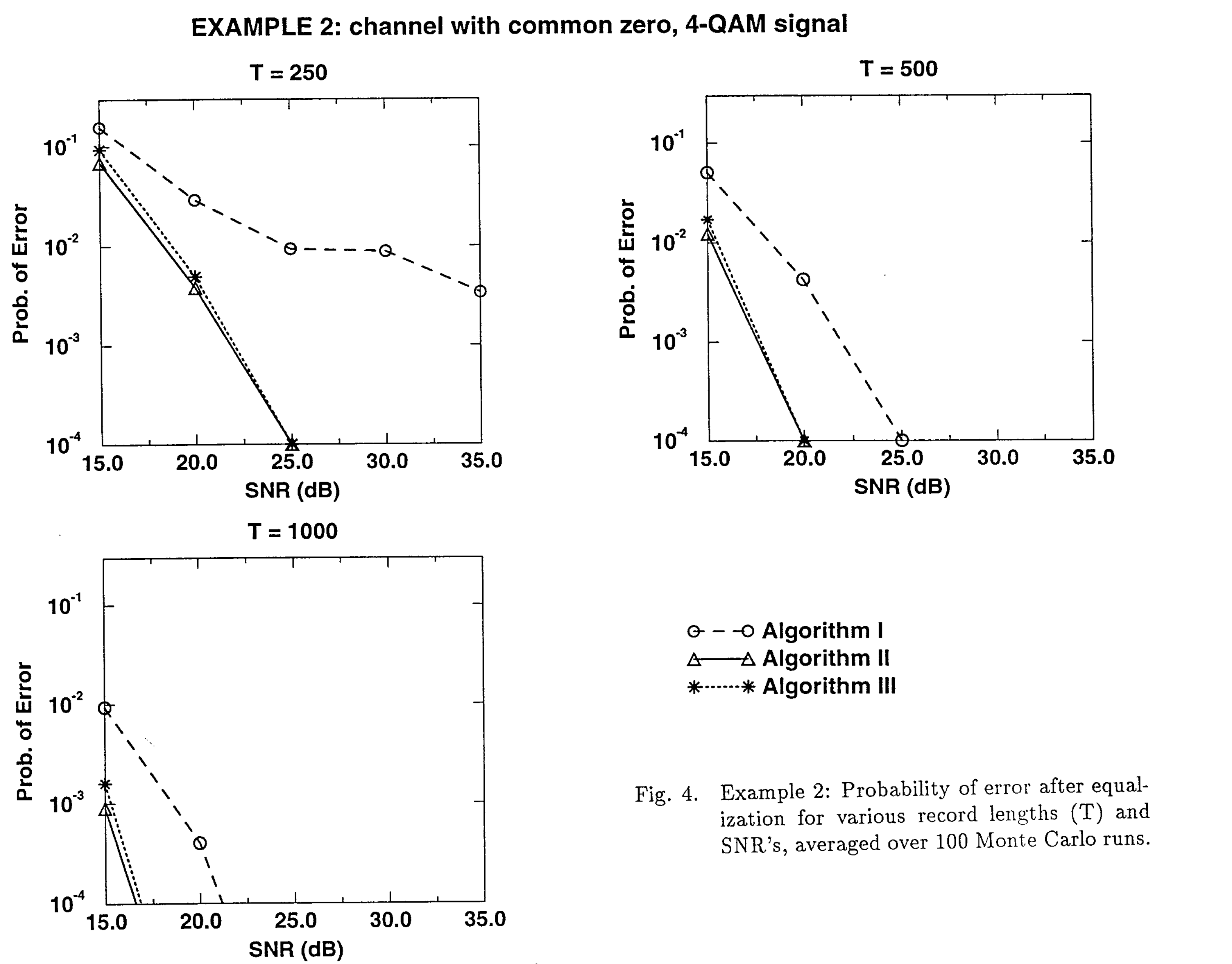


Fig. 4. Example 2: Probability of error after equalization for various record lengths (T) and SNR's, averaged over 100 Monte Carlo runs.

[4]

Multistep Linear Predictors-Based Blind Equalization Of FIR/IIR Single-Input Multiple-Output Channels With Common Zeros [1]

Jitendra K. Tugnait

Department of Electrical Engineering

Auburn University, Auburn, AL 36849, USA

Tel.: (334)844-1846 FAX: (334)844-1809

Email: tugnait@eng.auburn.edu

Abstract

The problem of blind equalization of SIMO (single-input multiple-output) communications channels is considered using only the second-order statistics of the data. Such models arise when a single receiver data is fractionally sampled (assuming that there is excess bandwidth), or when an antenna array is used with or without fractional sampling. We extend the multistep linear prediction approach to infinite impulse response (IIR) channels as well as to the case where the "subchannel" transfer functions have common zeros. In past this approach has been confined to finite impulse response (FIR) channels with no common subchannel zeros. We focus on direct design of finite-length MMSE (minimum mean-square error) blind equalizers. Knowledge of the nature of the underlying model (FIR or IIR) or the model order is not required. Our approach works when the "subchannel" transfer functions have common zeros so long as the common zeros are minimum-phase zeros. Illustrative simulation examples are provided.

SP EDICS : 2.8.1

[1]This work was supported by the Office of Naval Research under Grant N00014-97-1-0822.

1 Introduction

Consider a discrete-time SIMO (single-input multiple-output) system with N outputs and one input. The i-th component of the output at time k is given by

$$y_i(k) = \mathcal{F}_i(z)w(k) + n_i(k), \quad i = 1, 2, \cdots, N, \tag{1-1}$$

$$\Longrightarrow \quad \mathbf{y}(k) = \mathcal{F}(z)w(k) + \mathbf{n}(k) = \mathbf{s}(k) + \mathbf{n}(k), \tag{1-2}$$

where $\mathbf{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathbf{s}(k)$ and $\mathbf{n}(k)$, and z is the $\mathcal{Z}-$transform variable as well as the backward-shift operator (i.e., $z^{-1}w(k) = w(k-1)$, etc.). The sequence $w(k)$ is the (single) input at sampling time k, $y_i(k)$ is the i-th noisy output, $s_i(k)$ is the i-th noise-free output, $n_i(k)$ is the additive measurement noise, and

$$F_i(z) := \sum_{l=0}^{\infty} f_i(l) z^{-l} \tag{1-3}$$

is the scalar transfer function with $w(k)$ as the input and $y_i(k)$ as the output; it represents the $i-$th subchannel. We allow all of the above variables to be complex-valued. The overall transfer function is denoted by the $N \times 1$ $\mathcal{F}(z)$ with its $i-$th element as $\mathcal{F}_i(z)$. We have

$$\mathcal{F}(z) = \sum_{i=0}^{\infty} \mathbf{F}_i z^{-i}. \tag{1-4}$$

Such models arise in several useful baseband-equivalent digital communications and other applications. A case of some interest is that of fractionally-spaced samples of a single baseband received signal leading to a SIMO model [1],[4],[8]. Alternatively, a similar model can be derived when we have a single signal impinging upon an antenna array with N elements [5]. A similar model arises if we have an antenna array coupled with fractional sampling at each array-element [5].

In these applications one of the objectives is to recover the inputs $w(k)$ given the noisy measurements but not given the knowledge of the system transfer function. Recently there has been much interest in solving this problem using only (or at least, to the maximum extent possible) the second-order statistics (SOS) of the data (see [1], [3]-[5], [8]-[14] and references therein). The solution is closely tied to existence of an FIR (finite impulse response) inverse to the system transfer function [1], [3]-[5], [8]-[14]. An overwhelming number of papers (see [4],[5],[9]-[12] and references

therein) have concentrated on a two-step procedure: first estimate the channel impulse response (IR) and then design an equalizer using the estimated channel. A fundamental restriction in these works is that the channel is FIR with no common zeros among the various subchannels. A few (see [1]and [13], e.g.) have proposed direct design of the equalizer bypassing channel estimation. Still they assume FIR channels with no common zeros.

In this paper we allow IIR (infinite impulse response) channels (which are finitely parametrized). We will also allow common zeros so long as they are minimum-phase (i.e., they lie inside the unit circle). Finally, in the presence of nonminimum-phase common zeros, our proposed approach equalizes the spectrally-equivalent minimum-phase counterpart of $\mathcal{F}(z)$; it does not "fall apart" unlike quite a few existing approaches. Our proposed approach is inspired by that of [10] and [12] which have been derived and analyzed therein only for FIR channels with no common zeros. The basis for the proposed approach is multistep linear prediction. A one-step linear prediction-based approach was first proposed in [8] and later expanded upon in [9] and [14]. Unlike the subspace-based methods of [4], [5], [11] and others (see also [3] and references therein), the linear prediction (LP) based approach of [8], [9] and [14] turns out to be rather insensitive to the order of the underlying FIR channel (so long as one overfits). More recently, it has been pointed out in [10] and [12] that the LP-based approach can be further significantly improved by utilizing some additional information not exploited by LP. Although [10] and [12] derive their algorithms in a quite a different manner, their final algorithms are essentially the same. In this paper we will follow the approach of [12] which is based upon multistep linear prediction. As noted earlier, unlike [12] we allow IIR channels and common zeros.

Two approaches are discussed in this paper for designing a blind MMSE (minimum mean-square error) linear equalizer of a specified length and delay. The approaches do not require the knowledge of the underlying system model orders or IR length, nor do they require the knowledge of the nature of the model (FIR or IIR). Algorithm I is novel and is inspired by [10] and [12] whereas Algorithm II is a straightforward extension of [9] and [14], and it was first proposed in [18]. Note that the prediction error methods of [8], [9] and [14] apply to the problem under consideration with some straightforward extensions/modifications (as we discuss in Sec. 3.4.2). Interestingly, [8], [9] and [14]

derive their results under the assumption of FIR channels with no common zeros. Although our emphasis is on MMSE equalization, estimation of a leading part of the underlying channel IR is an essential part of this paper. For MMSE equalization with a given delay d, it is sufficient to estimate the channel IR at first $d+1$ samples, which is what is done in this paper. Clearly, if channel IR estimation is the objective, then one can pick a 'large' value of d.

The paper is organized as follows. Precise model assumptions and some background results used later in the paper are stated and developed in Sec. 2. IIR channels with no common subchannel zeros are considered in Sec. 3 where the proposed algorithm of this paper is developed. Under the assumptions of Sec. 3, finite length inverses and finite-length multistep linear predictors exist. In Sec. 4 we allow common subchannel zeros. Here ideally we need infinite length inverses and multistep linear predictors. Three computer simulation examples involving a 4-QAM signal are presented in Sec. 5 to illustrate the performance of the proposed approach and compare it with that of the linear prediction approach.

2 Model Assumptions and Preliminaries: No Common Zeros

In this section we consider precise model assumptions and some background results used later in the paper. In Secs. 2 and 3 we focus on systems with no common subchannel zeros. The case of common zeros is discussed in Sec. 4.

Assume the following:

(H1) $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ where $\mathcal{A}(z) = 1 + \sum_{i=1}^{n_a} a_i z^{-i}$ is 1×1, $\mathcal{B}(z) = \sum_{i=0}^{n_b} \mathbf{B}_i z^{-i}$ is $N \times 1$ and $N > 1$.

(H2) Rank$\{\mathcal{B}(z)\} = 1\ \forall z$ including $z = \infty$ but excluding $z = 0$, i.e., $\mathcal{B}(z)$ is irreducible [7, Sec. 6.3].

(H3) $\mathcal{A}(z) \neq 0$ for $|z| \geq 1$.

(H4) $\{w(k)\}$ is zero-mean, white. Take $E\{|w(k)|^2\} = 1$ by absorbing any non-identity correlation of $w(k)$ into $\mathcal{F}(z)$.

(H5) $\{\mathbf{n}(k)\}$ is zero-mean with $E\{\mathbf{n}(k+\tau)\mathbf{n}^H(k)\} = \sigma_n^2 I_{N\times N}$ where $I_{N\times N}$ is the $N \times N$ identity matrix and the superscript H is the Hermitian operator (complex conjugate transpose).

Assumption **(H2)** is equivalent to stating that the various subchannels $F_i(z)$ have no common zeros. It has been shown in [6] (using some results from [2]) that under **(H1)-(H3)** there exists a finite degree left-inverse (not necessarily unique) of $\mathcal{F}(z)$:

$$\mathcal{G}(z)\mathcal{F}(z) = 1 \tag{2-1}$$

where $\mathcal{G}(z)$ is $1 \times N$ given by

$$\mathcal{G}(z) = \sum_{l=0}^{L_e} \mathbf{G}_l z^{-l} \quad \text{for any} \quad L_e \geq n_a + n_b - 1. \tag{2-2}$$

Remark 1: The left-inverse $\mathcal{G}(z)$ of $\mathcal{F}(z)$ consists of two parts: $\mathcal{G}(z) = \mathcal{G}_B(z)\mathcal{A}(z)$ where $\mathcal{G}_B(z)\mathcal{B}(z) = 1$ so that $\mathcal{G}(z)\mathcal{F}(z) = \mathcal{G}_B(z)\mathcal{A}(z)\mathcal{A}^{-1}(z)\mathcal{B}(z) = \mathcal{G}_B(z)\mathcal{B}(z) = 1$. Finite length left-inverses of FIR SIMO channels have been subject of intense research activities [4]-[6],[8]-[13]. Left-inverses to MIMO IIR/FIR channels have been considered in [6]. It appears that the results of [6] pertaining to MIMO models are the sharpest to date. Finally, it is important to stress that [4], [5] and [8]-[13] do not allow IIR channels, or subchannels having common zeros, in their problem formulation unlike this contribution. □

2.1 MMSE Equalizer with Delay d

We wish to design an MMSE (minimum mean-square error) linear equalizer of a specified length. It is not too hard to establish (using the orthogonality principle [16], for example) that the MMSE equalizer of length $L_e + 1$ to estimate $w(k-d)$ ($d \geq 0$) based upon $\mathbf{y}(n)$, $n = k, k-1, \cdots, k-L_e$, satisfies

$$\begin{bmatrix} \overline{\mathbf{G}}_{d,0} & \overline{\mathbf{G}}_{d,1} & \cdots & \overline{\mathbf{G}}_{d,L_e} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_d^H & \mathbf{F}_{d-1}^H & \cdots & \mathbf{F}_0^H & 0 & \cdots & 0 \end{bmatrix} \mathcal{R}_{yyL_e}^{-1} \tag{2-3}$$

where $\mathcal{R}_{yyL_e}$ is a $[N(L_e+1)]\times[N(L_e+1)]$ matrix with its $ij-$th block-element given by $\mathbf{R}_{yy}(j-i) := E\{\mathbf{y}(k+j-i)\mathbf{y}^H(k)\}$. The equalized output is given by

$$\widehat{w}(k-d) = \sum_{i=0}^{L_e} \overline{\mathbf{G}}_{d,i}\mathbf{y}(k-i). \tag{2-4}$$

Clearly one can obtain a consistent estimate of $\mathcal{R}_{yyL_e}$ from the given data. It remains to estimate $\mathbf{F}_l$'s to complete the design. This is where the multistep predictor approach turns out to be useful.

3 Partial Channel Identification Using Multistep Predictors

3.1 FIR Multistep Linear Predictors

By (1-2) and (**H1**), it follows that

$$\mathrm{s}(k) = -\sum_{i=1}^{n_a} a_i \mathrm{s}(k-i) + \sum_{i=0}^{n_b} \mathbf{B}_i w(k-i). \tag{3-1}$$

It then follows from (3-1) that

$$\begin{aligned}\mathrm{s}(k) &= -\sum_{i=2}^{n_a} a_i \mathrm{s}(k-i) - a_1 \left[-\sum_{i=1}^{n_a} a_i \mathrm{s}(k-1-i) + \sum_{i=0}^{n_b} \mathbf{B}_i w(k-1-i) \right] + \sum_{i=0}^{n_b} \mathbf{B}_i w(k-i) \\ &= -\sum_{i=2}^{n_a+1} a_i^{(2)} \mathrm{s}(k-i) + \sum_{i=0}^{n_b+1} \mathbf{B}_i^{(2)} w(k-i). \end{aligned} \tag{3-2}$$

for some appropriate choices of the parameters $a_i^{(2)}$s and $\mathbf{B}_i^{(2)}$s. Now substitute for $\mathrm{s}(k-2)$ using (3-1) in (3-3), and continuing this way, we have, in general, for appropriate choices of $a_i^{(l)}$s and $\mathbf{B}_i^{(l)}$s ($l \geq 1$)

$$\mathrm{s}(k) = -\sum_{i=l}^{n_a+l-1} a_i^{(l)} \mathrm{s}(k-i) + \sum_{i=0}^{n_b+l-1} \mathbf{B}_i^{(l)} w(k-i). \tag{3-3}$$

Both (3-1) and (3-3) represent the same signal/system and therefore, they must have the same impulse response. By (1-4), (**H1**), (3-1) and (3-3), it then follows that

$$\mathbf{B}_i^{(l)} = \mathbf{F}_i \quad \text{for} \quad 0 \leq i \leq l-1. \tag{3-4}$$

Let us rewrite (3-3) as

$$\mathrm{s}(k) = \mathrm{e}(k|k-l) + \widehat{\mathrm{s}}(k|k-l) \tag{3-5}$$

where

$$\mathrm{e}(k|k-l) := \sum_{i=0}^{l-1} \mathbf{B}_i^{(l)} w(k-i) = \sum_{i=0}^{l-1} \mathbf{F}_i w(k-i) \tag{3-6}$$

and

$$\widehat{\mathrm{s}}(k|k-l) := -\sum_{i=l}^{n_a+l-1} a_i^{(l)}\mathrm{s}(k-i) + \sum_{i=l}^{n_b+l-1} \mathrm{B}_i^{(l)} w(k-i). \tag{3-7}$$

We first need some notations and definitions.

Notations and Definitions: Consider the Hilbert space $\mathcal{H}$ of square integrable complex random variables on a common probability space endowed with the inner product (for scaler complex random variables x_1 and x_2) $< x_1, x_2 >= E\{x_1 x_2^*\}$ where the superscript $*$ denotes complex conjugation (see [15]). Let $Sp\{x_i \in I\}$ denote the subspace of $\mathcal{H}$ generated by the random variables/vectors in the set $\{x_i \in I\}$. Let $H_k(\mathrm{s})$ denote the subspace generated by the past of s up to time k

$$H_k(\mathrm{s}) := Sp\{\ s_i(k-m),\ \ i = 1, 2, \cdots, N;\ \ m = 0, 1, \cdots\} \tag{3-8}$$

and let $H_{k-1,L}(\mathrm{s})$ denote the subspace spanned by a finite past of s

$$H_{k-1,L}(\mathrm{s}) := Sp\{\ s_i(k-m),\ \ i = 1, 2, \cdots, N;\ \ m = 1, 2, \cdots, L\}. \tag{3-9}$$

Let $(\mathrm{s}(k)|H_{k-1}(\mathrm{s}))$ denote the orthogonal projection of $\mathrm{s}(k)$ onto the subspace $H_{k-1}(\mathrm{s})$ [15]. $\square$

Theorem 1. Under (**H1**)-(**H4**) and for $l = 1, 2, \cdots$, $\{\mathrm{s}(k)\}$ can be decomposed as in (3-5) such that

$$E\{\mathrm{e}(k|k-l)\mathrm{s}^H(k-m)\} \ = \ 0 \quad \forall m \geq l, \tag{3-10}$$

$$\widehat{\mathrm{s}}(k|k-l) \ = \ (\mathrm{s}(k)|H_{k-l}(\mathrm{s}))\,, \tag{3-11}$$

$$\widehat{\mathrm{s}}(k|k-l) \ \in \ H_{k-l,n_a+n_b+l-1}(\mathrm{s}) \tag{3-12}$$

and

$$\widehat{\mathrm{s}}(k|k-l) \ = \ (\mathrm{s}(k)|H_{k-l,n_a+n_b+l-1}(\mathrm{s}))\,. \tag{3-13}$$

The decomposition (3-5) is unique. $\bullet$

Proof: By (1-2), (1-4), (**H1**) and (**H3**), we have

$$\mathrm{s}(k) \ = \ \sum_{i=0}^{\infty} \mathbf{F}_i w(k-i). \tag{3-14}$$

By (2-1), (2-2) and (1-2), it follows that

$$\sum_{i=0}^{L_e} \mathbf{G}_i \mathrm{s}(k-i) \;=\; w(k). \tag{3-15}$$

Substituting for $w(k)$ from (3-15) in (3-7), it follows that

$$\widehat{\mathrm{s}}(k|k-l) \;\in\; H_{k-l}(\mathrm{s}). \tag{3-16}$$

By (3-14) and (**H4**), we have

$$E\{w(k)\mathrm{s}^H(k-m)\} \;=\; 0 \quad \forall\, m>0. \tag{3-17}$$

Therefore, using (3-6) and (3-17), it follows that (3-10) is true. By (3-5), (3-10), (3-16) and the orthogonal projection theorem [15], it follows that (3-11) is true (as the "error" $\mathbf{e}(k|k-l)$ is orthogonal to the data $\mathrm{s}(k-m)$ $(m \geq l)$, hence to the subspace $H_{k-l}(\mathrm{s})$).

It remains to establish (3-12) and (3-13). Define

$$\mathrm{x}(k) \;:=\; \mathcal{A}(z)\mathrm{s}(k) \;=\; \mathcal{B}(z)w(k). \tag{3-18}$$

By the proof of Theorem 1 in [14] (and with obvious changes in notation of [14]), using (3-18) (i.e. $\mathrm{x}(k)=\mathcal{B}(z)w(k)$) and (**H2**), we have

$$H_{k,M}(\mathrm{x}) \;=\; H_{k,M+n_b}(w) \quad \forall\, M \geq n_b - 1. \tag{3-19}$$

It follows from (3-18) (i.e. $\mathrm{x}(k)=\mathcal{A}(z)\mathrm{s}(k)$) and (**H1**) that

$$H_{k,M}(\mathrm{x}) \;\subset\; H_{k,M+n_a}(\mathrm{s}) \quad \forall\, M \geq 0. \tag{3-20}$$

Therefore, (3-19) and (3-20) lead to

$$H_{k,M+n_b}(w) \;\subset\; H_{k,M+n_a}(\mathrm{s}) \quad \forall\, M \geq n_b - 1, \tag{3-21}$$

and in general, we have for any integer l

$$H_{k-l,M+n_b+l}(w) \;\subset\; H_{k-l,M+n_a+l}(\mathrm{s}) \quad \forall\, M \geq n_b - 1. \tag{3-22}$$

It therefore follows from (3-7) and (3-22) that

$$\widehat{\mathrm{s}}(k|k-l) \;\in\; H_{k-l,n_a+M+l}(\mathrm{s}) \quad \forall\, M \geq n_b - 1. \tag{3-23}$$

If we pick $M = n_b - 1$ in (3-23), we obtain (3-12). Finally, (3-13) follows from (3-5), (3-10), (3-12) and the orthogonal projection theorem [15].

Uniqueness of the decomposition (3-5) is a consequence of the orthogonal projection theorem [15]. Suppose that there exists some other decomposition

$$\mathrm{s}(k) = \widetilde{\mathrm{e}}(k|k-l) + \widehat{\widehat{\mathrm{s}}}(k|k-l) \tag{3-24}$$

such that

$$E\{\widetilde{\mathrm{e}}(k|k-l)\mathrm{s}^H(k-m)\} \;=\; 0 \quad \forall m \geq l \tag{3-25}$$

and

$$\widehat{\widehat{\mathrm{s}}}(k|k-l) \;\in\; H_{k-l}(\mathrm{s}). \tag{3-26}$$

Then the orthogonal projection theorem [15] implies that

$$E\{||\widehat{\mathrm{s}}(k|k-l) - \widehat{\widehat{\mathrm{s}}}(k|k-l)||^2\} \;=\; 0 \tag{3-27}$$

and

$$E\{||\mathrm{e}(k|k-l) - \widetilde{\mathrm{e}}(k|k-l)||^2\} \;=\; 0. \tag{3-28}$$

This completes the proof of Theorem 1. $\square$

Remark 2: When $\mathcal{A}(z) = 1$ (i.e. the channel is FIR), the results of Theorem 1 hold true with $n_a = 0$. Note that we may write $\mathcal{A}(z) = \sum_{i=0}^{n_a} a_i z^{-i}$ with $a_0 := 1$. $\square$

It follows from Theorem 1 that

$$\widehat{\mathrm{s}}(k|k-l) \;=\; \sum_{i=l}^{L_l} \mathrm{A}_i^{(l)}\mathrm{s}(k-i) \quad \text{where} \quad L_l \geq n_a + n_b + l - 1, \tag{3-29}$$

for some $N \times N$ matrices $\mathrm{A}_i^{(l)}$s. By (3-5) and (3-10) (recall also the orthogonal projection theorem), we have

$$\widehat{\mathrm{s}}(k|k-l) \;=\; \arg\left\{\min_{\mathrm{x}(k)\in H_{k-l}(\mathrm{s})} E\{||\mathrm{s}(k) - \mathrm{x}(k)||^2\right\}. \tag{3-30}$$

Therefore, $\widehat{\mathrm{s}}(k|k-l)$ is the $l-$step (ahead) linear predictor of $\mathrm{s}(k)$ given $\{\mathrm{s}(m),\ m \leq k-l\}$. By (3-13) it is also the $l-$step (ahead) linear predictor of $\mathrm{s}(k)$ given $\{\mathrm{s}(m),\ k - L_l \leq m \leq k-l\}$.

Using (3-5) and (3-29) we have

$$s(k) = \sum_{i=l}^{L_l} \mathbf{A}_i^{(l)} s(k-i) + \mathbf{e}(k|k-l). \tag{3-31}$$

By (3-10) and (3-31), for $m \geq l$,

$$E\{s(k)s^H(k-m)\} = \sum_{i=l}^{L_l} \mathbf{A}_i^{(l)} E\{s(k-i)s^H(k-m)\}. \tag{3-32}$$

By the orthogonal projection theorem and (3-13), it is sufficient to consider (3-32) for $m = l, l+1, \cdots, L_l$ in order to solve for $\mathbf{A}_i^{(l)}$s. Using these values of m in (3-32) we may write

$$\left[\begin{array}{cccc} \mathbf{A}_l^{(l)} & \mathbf{A}_{l+1}^{(l)} & \cdots & \mathbf{A}_{L_l}^{(l)} \end{array} \right] \mathcal{R}_{ss(L_l-l)} = \left[\begin{array}{cccc} \mathbf{R}_{ss}(l) & \mathbf{R}_{ss}(l+1) & \cdots & \mathbf{R}_{ss}(L_l) \end{array} \right] \tag{3-33}$$

where $\mathcal{R}_{ssM}$ denotes a $[N(M+1)] \times [N(M+1)]$ matrix with its ij-th block element as $\mathbf{R}_{ss}(j-i)$ $= E\{s(k+j-i)s^H(k)\}$. Note that $\mathcal{R}_{ss(L_l-l)}$ is not necessarily full rank, therefore, the coefficients $\mathbf{A}_i^{(l)}$s are not necessarily unique. (Note that the orthogonal projection theorem implies uniqueness in the sense of (3-27), it does not necessarily imply the uniqueness of the coefficients in (3-31).) A minimum norm solution to (3-33) may be obtained as [17]

$$\left[\begin{array}{cccc} \mathbf{A}_l^{(l)} & \mathbf{A}_{l+1}^{(l)} & \cdots & \mathbf{A}_{L_l}^{(l)} \end{array} \right] = \left[\begin{array}{cccc} \mathbf{R}_{ss}(l) & \mathbf{R}_{ss}(l+1) & \cdots & \mathbf{R}_{ss}(L_l) \end{array} \right] \mathcal{R}_{ss(L_l-l)}^{\#} \tag{3-34}$$

where the superscript # denotes the pseudoinverse.

3.2 Estimation of Noise Variance

Consider the case of $l = 1$ (one-step prediction). By (3-6) and (3-31) we have

$$s(k) = \sum_{i=1}^{L_1} \mathbf{A}_i^{(1)} s(k-i) + \mathbf{F}_0 w(k). \tag{3-35}$$

If $L_1 > n_a + n_b$, then $A_i^{(1)} = 0$ for $i > n_a + n_b$ by virtue of (3-13).

Lemma 1. Under **(H1)**-**(H4)**, $\rho(\mathcal{R}_{ssL_1}) \leq NL_1 + 1$ for $L_1 \geq n_a + n_b$ where $\rho(A)$ denotes the rank of A. •

Proof: It follows from (3-35) that

$$\left[\begin{array}{cccccc} I_{N\times N} & -\mathbf{A}_1^{(1)} & \cdots & -\mathbf{A}_{n_a+n_b}^{(1)} & 0 & \cdots & 0 \end{array} \right] \mathcal{R}_{ssL_1} = \left[\begin{array}{cccc} \mathbf{F}_0\mathbf{F}_0^H & 0 & \cdots & 0 \end{array} \right]. \tag{3-36}$$

Clearly

$$\rho\left(\left[\begin{array}{cccccccc} I_{N\times N} & -\mathbf{A}_1^{(1)} & \cdots & -\mathbf{A}_{n_a+n_b}^{(1)} & 0 & \cdots & 0 \end{array}\right]\right) = N \tag{3-37}$$

and

$$\rho\left(\left[\begin{array}{cccc} \mathbf{F}_0\mathbf{F}_0^H & 0 & \cdots & 0 \end{array}\right]\right) = 1. \tag{3-38}$$

Using (3-36)-(3-38) and Sylvester's inequality [7, p. 655], it follows that

$$\rho(\mathcal{R}_{ssL_1}) + N - N(L_1+1) \leq 1 \tag{3-39}$$

which yields the desired result. □

In a fashion similar to $\mathcal{R}_{ssM}$ in (3-33), let $\mathcal{R}_{yyM}$ denote a $[N(M+1)]\times[N(M+1)]$ matrix with its ij-th block element as $\mathbf{R}_{yy}(j-i) = E\{\mathbf{y}(k+j-i)\mathbf{y}^H(k)\}$; define similarly $\mathcal{R}_{nnM}$ pertaining to the additive noise. Carry out an eigendecomposition of $\mathcal{R}_{yyL_1}$. Then the smallest $N-1$ eigenvalues of $\mathcal{R}_{yyL_1}$ equal σ_n^2 because under **(H1)**-**(H4)**, $\rho(\mathcal{R}_{ssL_1}) \leq NL_1+1$ whereas under **(H5)**, $\rho(\mathcal{R}_{nnL_1}) = NL_1 + N = \rho(\mathcal{R}_{yyL_1})$. Thus a consistent estimate $\widehat{\sigma}_n^2$ of σ_n^2 is obtained by taking it as the average of the smallest $N-1$ eigenvalues of $\widehat{\mathcal{R}}_{yyL_1}$, the data-based consistent estimate of $\mathcal{R}_{yyL_1}$.

We will need the estimate of noise variance later to calculate $\mathcal{R}^{\#}_{ss(L_l-l)}$ in (3-34) for various $l \geq 1$. By (3-29), $L_l - l \geq n_a + n_b - 1$, independent of l. This suggests that we keep

$$L_l - l \;=\; \overline{L} \;\geq\; n_a + n_b - 1 \quad (\forall\, l). \tag{3-40}$$

Then under **(H4)** and **(H5)**,

$$\mathcal{R}_{ss\overline{L}} \;=\; \mathcal{R}_{yy\overline{L}} - \mathcal{R}_{nn\overline{L}} \;=\; \mathcal{R}_{yy\overline{L}} \;-\; \sigma_n^2 I_{(\overline{L}+1)\times(\overline{L}+1)}. \tag{3-41}$$

Thus, $\mathcal{R}_{ss\overline{L}}$ can be estimated from noisy data.

3.3 Partial Channel Identification

Recall that our main objective is to implement the MMSE linear equalizer with delay d as specified by (2-3) and (2-4). To this end, we need estimates of $\mathbf{F}_i$ for $i = 0, 1, \cdots, d$. We now show how (3-6) and (3-31) may be used for this purpose.

Let

$$\widetilde{L} = \overline{L} + d + 1 \tag{3-42}$$

where $\overline{L}$ is as in (3-40). Rewrite (3-31) as

$$\mathbf{e}(k|k-l) = \sum_{i=0}^{\widetilde{L}} \overline{\mathbf{A}}_i^{(l)} \mathbf{s}(k-i) \tag{3-43}$$

where

$$\overline{\mathbf{A}}_i^{(l)} = \begin{cases} I_{N\times N} & \text{for} \quad i = 0 \\ 0 & \text{for} \quad 1 \le i \le l-1 \\ -\mathbf{A}_i^{(l)} & \text{for} \quad l \le i \le \overline{L} + l \\ 0 & \text{for} \quad \overline{L} + l + 1 \le i \le \widetilde{L}\,. \end{cases} \tag{3-44}$$

By (3-40) $L_l = \overline{L} + l$, therefore, for each l, we estimate $\overline{L} + 1$ coefficients in (3-34). For $l \ge 2$, define

$$\begin{aligned} \overline{\mathbf{e}}_l(k) &:= \mathbf{e}(k|k-l) - \mathbf{e}(k|k-l+1) \\ &= \sum_{i=0}^{\widetilde{L}} \mathbf{D}_i^{(l,l-1)} \mathbf{s}(k-i) \end{aligned} \tag{3-45}$$

where

$$\mathbf{D}_i^{(l,l-1)} := \overline{\mathbf{A}}_i^{(l)} - \overline{\mathbf{A}}_i^{(l-1)}, \qquad i = 0, 1, \cdots, \widetilde{L}. \tag{3-46}$$

By (3-44), $\mathbf{D}_0^{(l,l-1)} = 0 \; \forall l \ge 2$.

Consider the $[N(d+1)]$−vector

$$\mathbf{E}(k) := \left[\overline{\mathbf{e}}_{d+1}^T(k+d) \vdots \overline{\mathbf{e}}_{d+1}^T(k+d-1) \vdots \cdots \vdots \overline{\mathbf{e}}_2^T(k+1) \vdots \mathbf{e}^T(k|k-1) \right]^T. \tag{3-47}$$

Using (3-43)-(3-47) we have

$$\mathbf{E}(k) = \mathcal{D} S(k) \tag{3-48}$$

where

$$S(k) := \left[\mathbf{s}^T(k+d-1) \vdots \mathbf{s}^T(k+d-2) \vdots \cdots \vdots \mathbf{s}^T(k-\widetilde{L}) \right]^T \tag{3-49}$$

is a $[N(\widetilde{L}+d)]$–column vector and

$$\mathcal{D} := \begin{bmatrix} \mathbf{D}_1^{(d+1,d)} & \mathbf{D}_2^{(d+1,d)} & \cdots & \mathbf{D}_{\widetilde{L}}^{(d+1,d)} & 0 & \cdots & \cdots & 0 & 0 \\ 0 & \mathbf{D}_1^{(d,d-1)} & \cdots & \mathbf{D}_{\widetilde{L}-1}^{(d,d-1)} & \mathbf{D}_{\widetilde{L}}^{(d,d-1)} & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \mathbf{D}_1^{(2,1)} & \cdots & \mathbf{D}_{\widetilde{L}}^{(2,1)} & 0 \\ 0 & 0 & \cdots & 0 & \cdots & \mathbf{A}_0^{(1)} & \cdots & \mathbf{A}_{\widetilde{L}-1}^{(1)} & \mathbf{A}_{\widetilde{L}}^{(1)} \end{bmatrix} \tag{3-50}$$

is a $[N(d+1)] \times [N(\widetilde{L}+d)]$ matrix. In (3-50) we have used the fact that $\mathbf{D}_0^{(l,l-1)} = 0 \; \forall l \geq 2$. Using (3-6), (3-45) and (3-47), we have

$$\mathbf{E}(k) = \begin{bmatrix} \mathbf{F}_d \\ \mathbf{F}_{d-1} \\ \vdots \\ \mathbf{F}_0 \end{bmatrix} w(k) =: \widetilde{\mathbf{F}} w(k). \tag{3-51}$$

By (3-48), (3-51) and **(H4)**, it follows that

$$\mathbf{R}_{EE}(0) = E\{\mathbf{E}(k)\mathbf{E}^H(k)\} = \widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H \tag{3-52}$$

$$= \mathcal{D}\mathcal{R}_{ss(\widetilde{L}+d-1)}\mathcal{D}^H. \tag{3-53}$$

Clearly $\rho(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H) = 1$. This suggests a method to estimate $\widetilde{\mathbf{F}}$. Calculate $\mathbf{R}_{EE}(0)$ as

$$\mathbf{R}_{EE}(0) = \mathcal{D}\left[\mathcal{R}_{yy(\widetilde{L}+d-1)} - \sigma_n^2 I_{(\widetilde{L}+d-1)\times(\widetilde{L}+d-1)}\right]\mathcal{D}^H. \tag{3-54}$$

Carry out an eigendecomposition of $\mathbf{R}_{EE}(0)$. The nonzero eigenvalue of $\mathbf{R}_{EE}(0)$ is equal to $\lambda = \|\widetilde{\mathbf{F}}\|^2$. Let the corresponding unit-norm eigenvector be denoted by Q_λ. Then

$$\widetilde{\mathbf{F}} = \alpha\sqrt{\lambda}Q_\lambda \tag{3-55}$$

for some α such that $|\alpha| = 1$. (We have the equality $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H = \lambda Q_\lambda Q_\lambda^H$ but not necessarily (3-55) with $\alpha = 1$.) In practice when the true values in (3-54) are replaced with their data-based (consistent) estimates, $\rho(\widehat{\mathbf{R}}_{EE}(0)) > 1$ where $\widehat{\mathbf{R}}_{EE}(0)$ denotes the estimate of $\mathbf{R}_{EE}(0)$. In this case we pick the largest eigenvalue as λ and the corresponding unit-norm eigenvector as Q_λ in order to implement (3-55). More details are provided in Sec. 3.4.

3.4 Practical Implementation

Given data $\mathbf{y}(k)$, $k = 1, 2, \cdots, T$. Pick the length $L_e + 1$ and delay d of the MMSE equalizer. (By (2-1) and (2-2) L_e should satisfy $L_e \geq n_a + n_b - 1$.) Let $\overline{L} = L_e$ in (3-40). The following steps are executed to implement a practical algorithm based on the earlier discussion in Secs. 2 and 3.1–3.3.

3.4.1 ALGORITHM I: Multistep Linear Predictors-Based Blind equalizer – MSLP

I.1 Estimate the correlation function of the measurements at lag m as

$$\widehat{\mathbf{R}}_{yy}(m) \;=\; \frac{1}{T}\sum_{k=1}^{T} \mathbf{y}(k+m)\mathbf{y}^H(k) \tag{3-56}$$

where we take $\mathbf{y}(k+m) = 0$ if $k+m < 1$ or $> T$. Define the $[N(\overline{L}+1)] \times [N(\overline{L}+1)]$ matrix $\widehat{\mathcal{R}}_{yy\overline{L}}$ with its $ij-$th block element as $\widehat{\mathbf{R}}_{yy}(j-i)$. Carry out an eigendecomposition of $\widehat{\mathcal{R}}_{yy\overline{L}}$. Let λ_i $(i = N\overline{L}+2, \cdots, N\overline{L}+N)$ denote the smallest $N-1$ eigenvalues of $\widehat{\mathcal{R}}_{yy\overline{L}}$. Estimate the noise variance σ_n^2 as

$$\widehat{\sigma}_n^2 \;=\; \frac{1}{N-1}\sum_{i=N\overline{L}+2}^{N\overline{L}+N} \lambda_i. \tag{3-57}$$

The signal correlation function at lag m is then estimated as

$$\widehat{\mathbf{R}}_{ss}(m) \;=\; \widehat{\mathbf{R}}_{yy}(m) \;-\; \widehat{\sigma}_n^2 I_{N\times N}\delta(m) \tag{3-58}$$

where $\delta(m)$ is the Kronecker delta function. Define the $[N(\overline{L}+1)] \times [N(\overline{L}+1)]$ signal correlation matrix estimate $\widehat{\mathcal{R}}_{ss\overline{L}}$ with its $ij-$th block element as $\widehat{\mathbf{R}}_{ss}(j-i)$. We will need (3-30) for $m = 0, 1, \cdots, \overline{L}+2d$ (see (3-42) and (3-54)).

I.2 Now we implement (3-34). First we need to calculate $\mathcal{R}^{\#}_{ss\overline{L}}$. Carry out a singular value decomposition of $\widehat{\mathcal{R}}_{ss\overline{L}}$ leading to $\widehat{\mathcal{R}}_{ss\overline{L}} = \mathbf{U}\Sigma\mathbf{V}^H$ where $\Sigma = diag\{s_i, i = 1, 2, \cdots, N\overline{L}+N\}$. The rank n_1 of $\widehat{\mathcal{R}}_{ss\overline{L}}$ is determined as the smallest n for which

$$q_n \;:=\; \sqrt{\frac{\sum_{i=n+1}^{N\overline{L}+N} s_i}{\sum_{i=1}^{N\overline{L}+N} s_i}} \leq \epsilon_1 \tag{3-59}$$

where $\epsilon_1 > 0$ is a small number. [For simulations presented in Sec. 5 we took $\epsilon_1 = 0.01$]. The desired pseudoinverse is then calculated as

$$\widehat{\mathcal{R}}^{\#}_{ss\overline{L}} = \mathbf{V}_1 \Sigma_1^{-1} \mathbf{U}_1^H \tag{3-60}$$

where $\Sigma_1 = diag\{s_i, i = 1, 2, \cdots, n_1\}$ and $\mathbf{U}_1$ and $\mathbf{V}_1$ are comprised of the left and the right (respectively) singular vectors corresponding to the singular values retained in Σ_1.

Estimate $\mathbf{A}_i^{(l)}$ for $i = l, l+1, \cdots, L_l$, $(L_l = \overline{L} + l)$ via (3-34) for $l = 1, 2, \cdots, d+1$:

$$\left[\begin{array}{cccc} \widehat{\mathbf{A}}_l^{(l)} & \widehat{\mathbf{A}}_{l+1}^{(l)} & \cdots & \widehat{\mathbf{A}}_{\overline{L}+l}^{(l)} \end{array} \right] = \left[\begin{array}{cccc} \widehat{\mathbf{R}}_{ss}(l) & \widehat{\mathbf{R}}_{ss}(l+1) & \cdots & \widehat{\mathbf{R}}_{ss}(\overline{L}+l) \end{array} \right] \widehat{\mathcal{R}}^{\#}_{ss\overline{L}}. \tag{3-61}$$

Following (3-44) set

$$\widehat{\overline{\mathbf{A}}}_i^{(l)} = \begin{cases} I_{N\times N} & \text{for} \quad i = 0 \\ 0 & \text{for} \quad 1 \le i \le l-1 \\ -\widehat{\mathbf{A}}_i^{(l)} & \text{for} \quad l \le i \le \overline{L}+l \\ 0 & \text{for} \quad \overline{L}+l+1 \le i \le \overline{L}+d+1 \end{cases} \tag{3-62}$$

I.3 Following (3-46) set

$$\widehat{\mathbf{D}}_i^{(l,l-1)} := \widehat{\overline{\mathbf{A}}}_i^{(l)} - \widehat{\overline{\mathbf{A}}}_i^{(l-1)}, \qquad i = 0, 1, \cdots, \overline{L}+d+1; \; l = 2, 3, \cdots, d+1, \tag{3-63}$$

and

$$\widehat{\mathcal{D}} = \text{ as in (3-50) with } \mathbf{D}_i^{(\cdot,\cdot)} \text{ replaced by } \widehat{\mathbf{D}}_i^{(\cdot,\cdot)}. \tag{3-64}$$

Define

$$\widehat{\mathbf{R}}_{EE}(0) = \widehat{\mathcal{D}} \widehat{\mathcal{R}}_{ss(\overline{L}+2d)} \widehat{\mathcal{D}}^H. \tag{3-65}$$

Carry out an eigendecomposition of $\widehat{\mathbf{R}}_{EE}(0)$ to calculate its largest eigenvalue $\widehat{\lambda}$ and the corresponding unit-norm eigenvector $\widehat{Q}_\lambda$. This yields the partial channel estimate up to a scale factor (recall (3-55)) as

$$\widehat{\overline{\mathbf{F}}} = \sqrt{\widehat{\lambda}} \widehat{Q}_\lambda. \tag{3-66}$$

I.4 Following (2-3), the MMSE equalizer with length L_e and delay d is calculated as

$$\left[\begin{array}{cccc} \widehat{\widehat{\mathbf{G}}}_{d,0} & \widehat{\widehat{\mathbf{G}}}_{d,1} & \cdots & \widehat{\widehat{\mathbf{G}}}_{d,L_e} \end{array} \right] = \widehat{\widehat{\mathbf{F}}}^H \widehat{\mathcal{R}}_{yyL_e}^{-1}. \tag{3-67}$$

Finally the equalized signal (up to a scale factor) is given by

$$\widehat{w}(k-d) = \sum_{i=0}^{L_e} \widehat{\widehat{\mathbf{G}}}_{d,i} \mathbf{y}(k-i). \tag{3-68}$$

3.4.2 ALGORITHM II: Linear Predictor-Based Blind equalizer – LP

Here we will use (2-3) with $\mathbf{F}_i$ ($i = 0, 1, \cdots, d$) estimated using the basic approach of [9] and [14] which utilizes one-step ahead linear predictors ($l = 1$). Although [9] and [14] derive all their results under the assumption of FIR channels with no common zeros, their results extend (with straightforward modifications) to models that satisfy (**H1**)-(**H5**) by virtue of Theorem 1. By (3-6) and (3-35), we have

$$w(k) = \|\mathbf{F}_0\|^{-2} \mathbf{F}_0^H \mathbf{e}(k|k-1). \tag{3-69}$$

From (3-14) and (**H4**), we have the relations

$$E\{w(k-l)\mathbf{s}^H(k)\} = \mathbf{F}_l^H \quad \text{for} \quad l \geq 0. \tag{3-70}$$

From (3-35) and (3-69), we have the relations

$$E\{w(k-l)\mathbf{s}^H(k)\} = \|\mathbf{F}_0\|^{-1} \mathbf{F}_0^H \left[\mathbf{R}_{ss}(-l) + \sum_{i=1}^{L_1} \mathbf{A}_i^{(1)} \mathbf{R}_{ss}(-l-i) \right]. \tag{3-71}$$

From (3-71) and (3-72) it follows that

$$\mathbf{F}_l^H = \|\mathbf{F}_0\|^{-1} \mathbf{F}_0^H \left[\mathbf{R}_{ss}^H(l) + \sum_{i=1}^{L_1} \mathbf{A}_i^{(1)} \mathbf{R}_{ss}^H(l+i) \right]. \tag{3-72}$$

Based upon the above discussion, [9] and [14], we have the following algorithm:

II.1 Repeat step I.1 of Algorithm I.

II.2 Execute step I.2 of Algorithm I only for $l = 1$. Calculate

$$\widehat{\mathbf{R}}_{ee}(0) := \widehat{\mathbf{R}}_{ss}(0) - \sum_{i=1}^{L_1} \widehat{\mathbf{A}}_i^{(1)} \widehat{\mathbf{R}}_{ss}(-i). \tag{3-73}$$

Carry out an eigendecomposition of $\widehat{\mathbf{R}}_{ee}(0)$ to calculate its largest eigenvalue $\widehat{\lambda}_0$ and the corresponding unit-norm eigenvector $\widehat{Q}_{\lambda_0}$. This yields the estimate (up to a scale factor) of $\mathbf{F}_0$ as

$$\widehat{\mathbf{F}}_0 = \sqrt{\widehat{\lambda}_0}\, \widehat{Q}_{\lambda_0}. \tag{3-74}$$

II.3 Estimate $\mathbf{F}_l^H$ up to a scale factor as

$$\widehat{\mathbf{F}}_l^H = [\widehat{\lambda}_0]^{-1} \widehat{\mathbf{F}}_0^H \left[\widehat{\mathbf{R}}_{ss}^H(l) + \sum_{i=1}^{L_1} \widehat{\mathbf{A}}_i^{(1)} \widehat{\mathbf{R}}_{ss}^H(l+i) \right], \quad l = 0, 1, \cdots, d. \tag{3-75}$$

II.4 The MMSE equalizer of length $L_e + 1$ and with delay d is calculated (up to a scale factor) as

$$\left[\begin{array}{cccc} \widehat{\widehat{\mathbf{G}}}_{d,0} & \widehat{\widehat{\mathbf{G}}}_{d,1} & \cdots & \widehat{\widehat{\mathbf{G}}}_{d,L_e} \end{array} \right] = \left[\begin{array}{ccccccc} \widehat{\mathbf{F}}_d^H & \widehat{\mathbf{F}}_{d-1}^H & \cdots & \widehat{\mathbf{F}}_0^H & 0 & \cdots & 0 \end{array} \right] \widehat{\mathcal{R}}_{yyL_e}^{-1}. \tag{3-76}$$

Finally, execute (3-68).

4 Blind Equalization: Common Zeros

Now we allow common subchannel zeros. In this case since ideally we need infinite length inverses and linear predoctors, the presented results hold true only approximately for finite length equalizers. Assume that **(H1)**-**(H5)** hold true.

4.1 Minimum-Phase Zeros

Here the SIMO transfer function is

$$\mathcal{F}(z) = \frac{\mathcal{B}_c(z)}{\mathcal{A}(z)} \mathcal{B}(z) = \mathcal{F}_1(z) \mathcal{B}_c(z) \tag{4-1}$$

where $\mathcal{F}_1(z)$ satisfies **(H1)**, $\mathcal{B}(z)$ satisfies **(H2)** and $\mathcal{B}_c(z)$ is a finite-degree scalar polynomial that collects all the common zeros of the subchannels. Assume that

(H6) Given model (4-1), $\mathcal{B}_c(z) \neq 0$ for $|z| \geq 1$.

Then while $\mathcal{F}_1(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ has a finite inverse, $\mathcal{B}_c^{-1}(z)$ is IIR though causal under **(H6)**.

By **(H6)** there exists a unique scalar polynomial $\mathcal{G}_c(z)$ such that

$$\mathcal{G}_c(z)\mathcal{B}_c(z) = 1 \quad \text{where} \quad \mathcal{G}_c(z) = \sum_{i=0}^{\infty} g_i z^{-i} \tag{4-2}$$

and

$$\sum_{i=0}^{\infty} |g_i|^2 < \infty. \tag{4-3}$$

Indeed, for some $0 < \alpha_1 < \infty$ and $0 < \beta_1 < 1$, we have

$$|g_i| \leq \alpha_1 \beta_1^{|i|} \quad \forall i. \tag{4-4}$$

Using (4-1)–(4-4), (2-1) and (2-2), it follows that there exists a $1 \times N$ polynomial $\mathcal{G}'(z)$ such that for some $0 < \alpha_2 < \infty$ and $0 < \beta_2 < 1$,

$$\mathcal{G}'(z) = \sum_{i=0}^{\infty} \mathbf{G}_i' z^{-i}, \tag{4-5}$$

$$\|\mathbf{G}_i'\| \leq \alpha_2 \beta_2^{|i|} \quad \forall i \tag{4-6}$$

and

$$\mathcal{G}'(z)\mathcal{F}(z) = 1. \tag{4-7}$$

By (1-4), **(H1)** and **(H3)**, for some $0 < \alpha_3 < \infty$ and $0 < \beta_3 < 1$, we have

$$\|\mathbf{F}_i\| \leq \alpha_3 \beta_3^{|i|} \quad \forall i. \tag{4-8}$$

Consider

$$\widehat{\mathrm{s}}(k|k-l) := \sum_{i=l}^{\infty} \mathbf{F}_i w(k-i). \tag{4-9}$$

By (4-5)–(4-7) and (1-2), it follows that

$$w(k) = \sum_{i=0}^{\infty} \mathbf{G}_i' \mathrm{s}(k-i). \tag{4-10}$$

Substituting for $w(k)$ from (4-10) in (4-9), it follows that

$$\widehat{\mathrm{s}}(k|k-l) = \sum_{i=l}^{\infty} \mathbf{F}_i \left(\sum_{m=0}^{\infty} \mathbf{G}'_m \mathrm{s}(k-i-m) \right) \tag{4-11}$$

$$= \sum_{n=l}^{\infty} \mathbf{C}_n \mathrm{s}(k-n) \tag{4-12}$$

where (recall that $\mathbf{F}_i = 0$ for $i < 0$)

$$\mathbf{C}_n = \sum_{m=0}^{\infty} \mathbf{F}_{n-m} \mathbf{G}'_m = \sum_{m=0}^{n} \mathbf{F}_{n-m} \mathbf{G}'_m. \tag{4-13}$$

It follows from (4-6), (4-8) and (4-13) that for some $0 < \alpha_4 < \infty$ and $0 < \beta_4 < 1$, we have

$$\|\mathbf{C}_i\| \leq \alpha_4 \beta_4^{|i|} \quad \forall i. \tag{4-14}$$

We now rewrite (3-14) as

$$\mathrm{s}(k) = \mathrm{e}(k|k-l) + \widehat{\mathrm{s}}(k|k-l) \tag{4-15}$$

where $\mathrm{e}(k|k-l)$ is as in (3-6) ,but $\widehat{\mathrm{s}}(k|k-l)$ is given by (4-12). We have the following result.

Theorem 2. Under (**H1**)-(**H4**) with $\mathcal{F}(z)$ in (**H1**) obeying (4-1), (**H6**) and for $l = 1, 2, \cdots$, $\{\mathrm{s}(k)\}$ can be decomposed as in (4-15) such that

$$E\{\mathrm{e}(k|k-l)\mathrm{s}^H(k-m)\} = 0 \quad \forall m \geq l, \tag{4-16}$$

and

$$\widehat{\mathrm{s}}(k|k-l) = (\mathrm{s}(k)|H_{k-l}(\mathrm{s})). \tag{4-17}$$

Furthermore, let

$$\widehat{\mathrm{s}}(k|k-l, k-M) := (\mathrm{s}(k)|H_{k-l,M}(\mathrm{s})) \tag{4-18}$$

and

$$\mathrm{e}(k|k-l, k-M) := \mathrm{s}(k) - \widehat{\mathrm{s}}(k|k-l, k-M). \tag{4-19}$$

Then

$$\lim_{M\to\infty} E\{\|\mathrm{e}(k|k-l) - \mathrm{e}(k|k-l, k-M)\|^2\} = 0 \quad \bullet \tag{4-20}$$

Proof: Eqns. (4-16) and (4-17) follow as in Theorem 1. We now turn to (4-20). It follows from (4-17), (4-18) and [15, Chapter 1, Lemma 3.1.b] that

$$\lim_{M\to\infty} E\{\|\widehat{\mathrm{s}}(k|k-l) - \widehat{\mathrm{s}}(k|k-l,k-M)\|^2\} \;=\; 0. \tag{4-21}$$

Then (4-20) is immediate via (4-15) and (4-19). We will provide an alternative, interesting proof. Consider

$$\begin{aligned} E\{\|\mathrm{e}(k|k-l) - \mathrm{e}(k|k-l,k-M)\|^2\} \;&=\; E\{\|\mathrm{e}(k|k-l)\|^2\} + E\{\|\mathrm{e}(k|k-l,k-M)\|^2\} \\ &-E\{\mathrm{e}^H(k|k-l)\mathrm{e}(k|k-l,k-M)\} - E\{\mathrm{e}^H(k|k-l,k-M)\mathrm{e}(k|k-l)\}. \end{aligned} \tag{4-22}$$

Using (4-15), (4-17)–(4-19) and the orthogonality principle, it follows that

$$E\{\mathrm{e}^H(k|k-l)\mathrm{e}(k|k-l,k-M)\} \;=\; E\{\mathrm{e}^H(k|k-l)\mathrm{s}(k)\} \;=\; E\{\|\mathrm{e}(k|k-l)\|^2\}. \tag{4-23}$$

Hence by (4-22) and (4-23), we have

$$E\{\|\mathrm{e}(k|k-l) - \mathrm{e}(k|k-l,k-M)\|^2\} = E\{\|\mathrm{e}(k|k-l,k-M)\|^2\} - E\{\|\mathrm{e}(k|k-l)\|^2\}. \tag{4-24}$$

Define

$$\widehat{\mathrm{s}}_M(k|k-l) \;:=\; \sum_{n=l}^{M} \mathbf{C}_n \mathrm{s}(k-n) \tag{4-25}$$

where $\mathbf{C}_n$s satisfy (4-13). Then

$$\begin{aligned} E\{\|\mathrm{e}(k|k-l,k-M)\|^2\} \;&\le\; E\{\|\mathrm{s}(k) - \widehat{\mathrm{s}}_M(k|k-l)\|^2\} \\ &=\; E\{\|\mathrm{s}(k) - \widehat{\mathrm{s}}(k|k-l)\|^2\} \;+\; E\{\|\widehat{\mathrm{s}}(k|k-l) - \widehat{\mathrm{s}}_M(k|k-l)\|^2\} \\ &=\; E\{\|\mathrm{e}(k|k-l)\|^2\} \;+\; E\{\|\widehat{\mathrm{s}}(k|k-l) - \widehat{\mathrm{s}}_M(k|k-l)\|^2\} \end{aligned} \tag{4-26}$$

where we have used the facts that (4-18) holds true, $\widehat{\mathrm{s}}_M(k|k-l) \in H_{k-l,M}(\mathrm{s})$ and $\mathrm{e}(k|k-l) = \mathrm{s}(k) - \widehat{\mathrm{s}}(k|k-l)$ is orthogonal to $H_{k-l}(\mathrm{s})$. By (4-24) and (4-26) we have

$$0 \;\le\; E\{\|\mathrm{e}(k|k-l) - \mathrm{e}(k|k-l,k-M)\|^2\} \;\le\; E\{\|\widehat{\mathrm{s}}(k|k-l) - \widehat{\mathrm{s}}_M(k|k-l)\|^2\}. \tag{4-27}$$

By (4-12) and (4-25), it follows that

$$\widehat{\mathrm{s}}(k|k-l) - \widehat{\mathrm{s}}_M(k|k-l) \;=\; \sum_{n=M+1}^{\infty} \mathbf{C}_n \mathrm{s}(k-n). \tag{4-28}$$

It then follows from (3-14), (4-14) and (4-28) that

$$\lim_{M\to\infty} E\{\|\widehat{\mathbf{s}}(k|k-l) - \widehat{\mathbf{s}}_M(k|k-l)\|^2\} \;=\; 0. \tag{4-29}$$

The desired result (4-20) then follows from (4-27) and (4-29). $\square$

Theorem 2 clearly implies that for M 'large enough' in (4-18), we can obtain $\mathbf{e}(k|k-l, k-M)$ close enough to $\mathbf{e}(k|k-l)$. Therefore, the approach of Sec. 3 becomes applicable to the current case. Note that validity of (2-3) and (2-4) is unaffected by **(H6)**. For a fixed d and L_e in (2-3) and (2-4), one needs to estimate $\mathbf{F}_i$ for $i = 0, 1, \cdots, d$. To this end, one should pick a 'large' $\overline{L}$ in (3-40) and (3-42), and unlike Sec. 3.4, it need not be equal to L_e, rather $\overline{L} > L_e$.

Remark 3: The alternative proof of (4-20) given above may be used to obtain an upper bound on the approximation error in (4-20) for finite M. By (4-27) and (4-28) we have

$$E\{\|\mathbf{e}(k|k-l) - \mathbf{e}(k|k-l, k-M)\|^2\} \;\le\; \mathrm{tr}\left\{ \sum_{n=M+1}^{\infty} \sum_{m=M+1}^{\infty} \mathbf{C}_n \mathbf{R}_{ss}(m-n)\mathbf{C}_m^H \right\}. \tag{4-30}$$

4.2 Arbitrary Zeros

In this case (4-1) is true but $\mathcal{B}_c(z)$ does not necessarily satisfy **(H6)**. We may rewrite (4-1) as

$$\mathcal{F}(z) \;=\; \overline{\mathcal{F}}(z)\mathcal{F}_{AP}(z) \tag{4-31}$$

where $\mathcal{F}_{AP}(z)$ is an allpass (rational) function such that

$$\mathcal{B}_c(z)\mathcal{B}_c(z^{-1}) \;=\; \mathcal{F}_{AP}(z)\overline{\mathcal{B}}_{MP}(z) \tag{4-32}$$

and $\overline{\mathcal{B}}_{MP}(z)$ is minimum-phase. Thus (within a scale factor) we have

$$\overline{\mathcal{F}}(z) \;=\; \frac{\overline{\mathcal{B}}_{MP}(z)}{\mathcal{A}(z)}\mathcal{B}(z). \tag{4-33}$$

We may rewrite (1-2) as

$$\mathbf{y}(k) \;=\; \overline{\mathcal{F}}(z)w'(k) + \mathbf{n}(k) \tag{4-34}$$

where

$$w'(k) \;:=\; \mathcal{F}_{AP}(z)w(k). \tag{4-35}$$

Clearly $w'(k)$ satisfies (**H4**). Hence, (4-33)-(4-35) satisfy the requirements of Sec. 4.1. Therefore, one can "approximately" recover $w'(k)$ from the given data by applying the algorithm of Sec. 3.4.

In order to recover $w(k)$ form $w'(k)$, one needs to exploit the higher-order statistics of $\{w'(k)\}$; see [2],[3] and references therein.

5 Simulation Examples

Here we consider three simulation examples to illustrate the proposed approaches. The first two examples are modified versions of the example from [10]. Example 1 consists of an ARMA model whose MA part is taken from [10]. Example 2 consists of an MA (FIR) model where we augment the FIR channel of [10] with a zero at 0.5 where this zero is common to all of the three subchannels. Finally, the channel model in Example 3 is exactly as in [10].

For computing $\widehat{\mathcal{R}}^{\#}_{ssL_e}$ in (3-60) via SVD, we picked $\epsilon_1 = 0.01$ in (3-59). Moreover, $\widehat{\mathcal{R}}^{-1}_{yyL_e}$ in (3-67) and (3-76) was also computed using SVD where all singular values smaller than $0.002\times$(largest singular value) were neglected. Thus, calculation of $\widehat{\mathcal{R}}^{-1}_{yyL_e}$ was regularized. The measurement SNR is defined as

$$\mathrm{SNR} \;=\; \frac{\sum_{i=1}^{N} E\{|s_i(k)|^2\}}{\sum_{i=1}^{N} E\{|n_i(k)|^2\}}.$$

The normalized MSE (i.e., MSE divided by $E\{|w(k)|^2\}$) and the probability of detection error (P_e) after equalization were taken as the two performance measures after averaging over 100 Monte Carlo runs. The equalized data were rotated and scaled before calculating the two performance measures. After designing the equalizers based on the given data record, the designed equalizer was applied to an independent record of length 3000 symbols in order to calculate normalized MSE and P_e. Therefore, the estimated P_e is not reliable below approximately 10^{-4}, hence, these values are not shown in Figs. 1–3.

5.1 Example 1: IIR Channel With No Common Zero

We have $N = 3$ in (1-2) with $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ where

$$\mathcal{A}(z) = (1 - 0.5z^{-1})I_{3\times 3} \tag{5-1}$$

and $\mathcal{B}(z)$ is 3×1 MA(6) obtained from [10] as follows. Consider a raised cosine pulse $p_6(t, 0.1)$ with a roll-off factor 0.1, truncated to a length of $6T_s$ (T_s = symbol duration). As in [10], a two-ray multipath channel with (effective) impulse response

$$h(t) = p_6(t, 0.1) - 0.7p_6(t - T_s/3, 0.1) \tag{5-2}$$

was sampled at intervals of $T_s/3$ (starting at $t = -3T_s$) to create the $\mathcal{B}(z)$ above. Transfer function $\mathcal{B}(z)$ satisfies (**H2**) [10], therefore, there exists a finite left inverse of length $L_e = 6$ (cf. Sec. 2).

The scalar input $w(k)$ is 4-QAM. An MMSE equalizer of length $L_e = 8$ (9 taps per subchannel, totaling 27 taps -- overfitting) was designed with a delay d =3 (arbitrarily selected just for illustration). The Algorithms I (MSLP) and II (LP) were applied for record lengths $T = 250$ and 500 symbols with varying SNR's. In order to apply MSLP, we picked $\overline{L} = L_e = 8$ ($> n_a + n_b - 1 = 1 + 6 - 1 = 6$) in (3-40). We picked $L_1 = L_e = 8$ for LP in (3-73) and (3-75). Fig. 1 shows the normalized MSE and the probability of error P_e. It is seen that the proposed design approach can handle IIR channels with little difficulty. Algorithm I (MSLP) performs the best.

5.2 Example 2: FIR Channel With Common Zero

Again we have $N = 3$ in (1-2) but with $\mathcal{F}(z) = \mathcal{B}_c(z)\mathcal{B}(z)$ where $\mathcal{B}(z)$ is as in Example 1 and $\mathcal{B}_c(z)$ is a scalar polynomial given by

$$\mathcal{B}_c(z) \ = \ 1 - 0.5z^{-1}. \tag{5-3}$$

Thus all three subchannels have a common zero at 0.5. The input $w(k)$ is 4-QAM as in Example 1. Note that in this example a finite left inverse and finite-length multistep predictors do not exist. First, as in Example 1, an MMSE equalizer of length $L_e = 8$ was designed with a delay d =3. The various design parameters for MSLP and LP ($\overline{L}$ and L_1) were as in Example 1. Fig. 2 shows the normalized MSE and P_e. We also tried a longer equalizer with $L_e = 12$ and $d = 3$. Furthermore, we picked $\overline{L} = L_e = 12$ for MSLP and $L_1 = L_e = 12$ for LP. The normalized MSE and P_e for this choice is shown in Fig. 3. It is seen from Figs. 2 and 3 that the proposed design approach can handle subchannels with common minimum-phase zeros with little difficulty, and that the approaches are not unduly sensitive to the choice of the various parameters involved. As in Example 1, Algorithm I performs the best.

5.3 Example 3: FIR Channel With No Common Zeros

This channel is exactly as in [10] with $N = 3$ in (1-2) and $\mathcal{F}(z) = \mathcal{B}(z)$ where $\mathcal{B}(z)$ is as in Example 1. As in Example 1, an MMSE equalizer of length $L_e = 8$ was designed with a delay d =3 and the design parameters for MSLP and LP were kept unchanged. Fig. 4 shows the normalized MSE and P_e. As in the earlier examples, MSLP outperforms LP.

6 Conclusions

Direct blind MMSE equalization of SIMO channels using only the second-order statistics of the data and the multistep linear prediction approach was considered. Such channels arise when antenna arrays are used or when signals with excess bandwidth are fractionally sampled or when both these scenarios are applicable. Unlike the past work on this problem [4],[5],[8]-[14], the proposed solution is applicable to IIR channels and to SIMO systems having common zeros among the various subchannels so long as the common zeros are minimum-phase. In case of nonminimum-phase zeros, we recover an allpass filtered version of the original input.

Computer simulation examples show that the multistep linear predictors-based MMSE equalizer outperforms the one-step linear predictor-based MMSE equalizer by a wide margin.

Future work includes performance analysis, adaptive implementation and extension to MIMO scenarios involving more than one information signals.

7 References

[1] G.B. Giannakis and S.D. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: direct and adaptive solutions," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 2277-2292, Sept. 1997.

[2] J.K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 268-271, Jan. 1997.

[3] Special Issue, *IEEE Transactions on Signal Processing*, vol. SP-45, Jan. 1997.

[4] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channels," *IEEE Trans. Information Theory*, vol. IT-40, pp. 340-349, March 1994.

[5] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 516-525, Feb. 1995.

[6] J.K. Tugnait, "FIR inverses to MIMO rational transfer functions with application to blind equalization," in *Proc. 30th Annual Asilomar Conf. Signals Systems Computers*, pp. 295-299, Pacific Grove, CA, Nov. 1996.

[7] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[8] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction," in *Proc. 1994 IEEE ICASSP*, pp. IV:585-588, Adelaide, Australia, May 1994.

[9] K. Abed-Meraim et al. "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. 1995 IEEE ICASSP*, pp. 1968-1971, Detroit, MI, May 9-12, 1995.

[10] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 3053-3061, Dec. 1997.

[11] K. Abed-Meraim et al. "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 42-55, Jan. 1997.

[12] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. 1997 ICASSP*, pp. 3621-3624, April 21-24, 1997.

[13] D. Gesbert, P. Duhamel and S. Mayrargue, "Blind multichannel adaptive MMSE equalization with controlled delay," in *Proc. Eighth IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 172-175, Corfu, Greece, June 24-26, 1996.

[14] K. Abed-Meraim et al. "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 694-705, March 1997.

[15] P.E. Caines, *Linear Stochastic Systems.* Wiley: New York, 1988.

[16] H.V. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag: New York, 1988.

[17] D.G. Luenberger, *Optimization by Vector Space Methods.* New York: John Wiley, 1969.

[18] B. Huang and J.K. Tugnait, "Blind equalization of I.I.R. single-input multiple-output channels with common zeros using second-order statistics," in *Proc. 1998 IEEE ICASSP*, Seattle, WA, May 12-15, 1998.

EXAMPLE 1: IIR channel, 9X3 taps

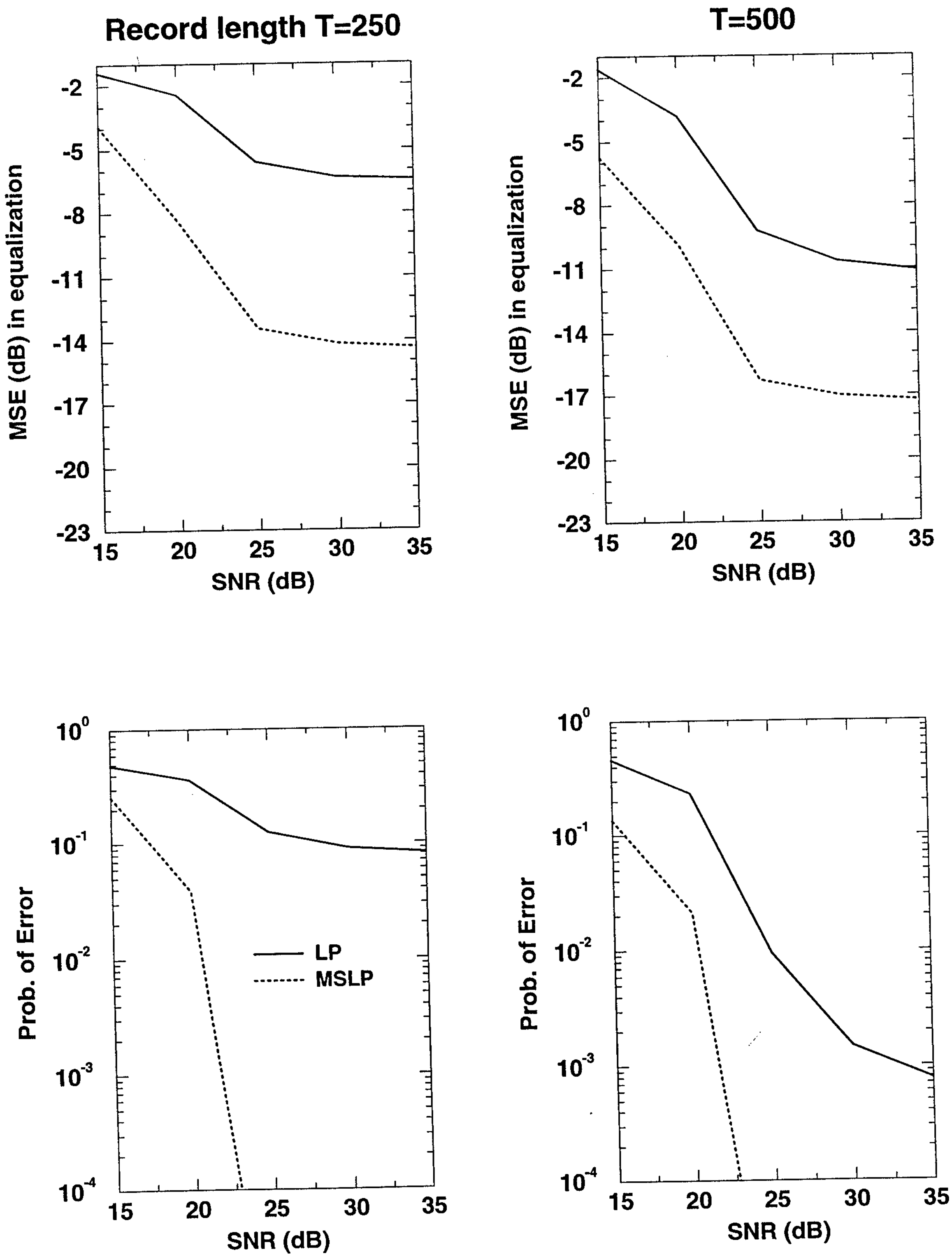


Fig. 1. Example 1: Probability of detection error and normalized MSE after equalization for various SNR's, averaged over 100 Monte Carlo runs. Left column: record length T=250 symbols; right column: record length T=500 symbols. Solid lines: Algorithm II (LP); dotted lines: Algorithm I (MSLP). Parameters: $\overline{L} = L_e = L_1 = 8$.

EXAMPLE 2: FIR channel with common zero, 9X3 taps

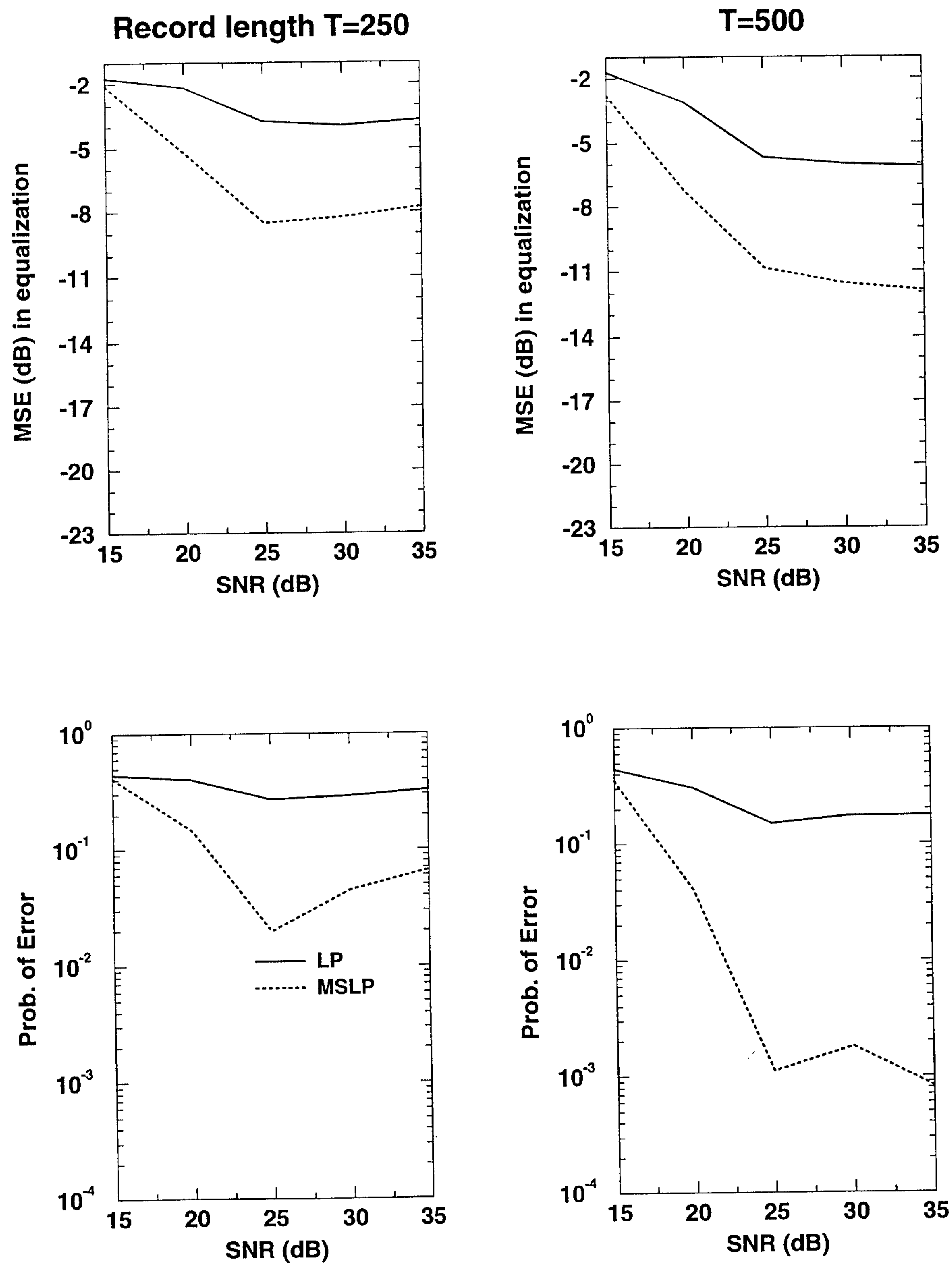


Fig. 2. Example 2: Probability of detection error and normalized MSE after equalization for various SNR's, averaged over 100 Monte Carlo runs. Left column: record length T=250 symbols; right column: record length T=500 symbols. Solid lines: Algorithm II (LP); dotted lines: Algorithm I (MSLP). Parameters: $\overline{L} = L_e = L_1 = 8$.

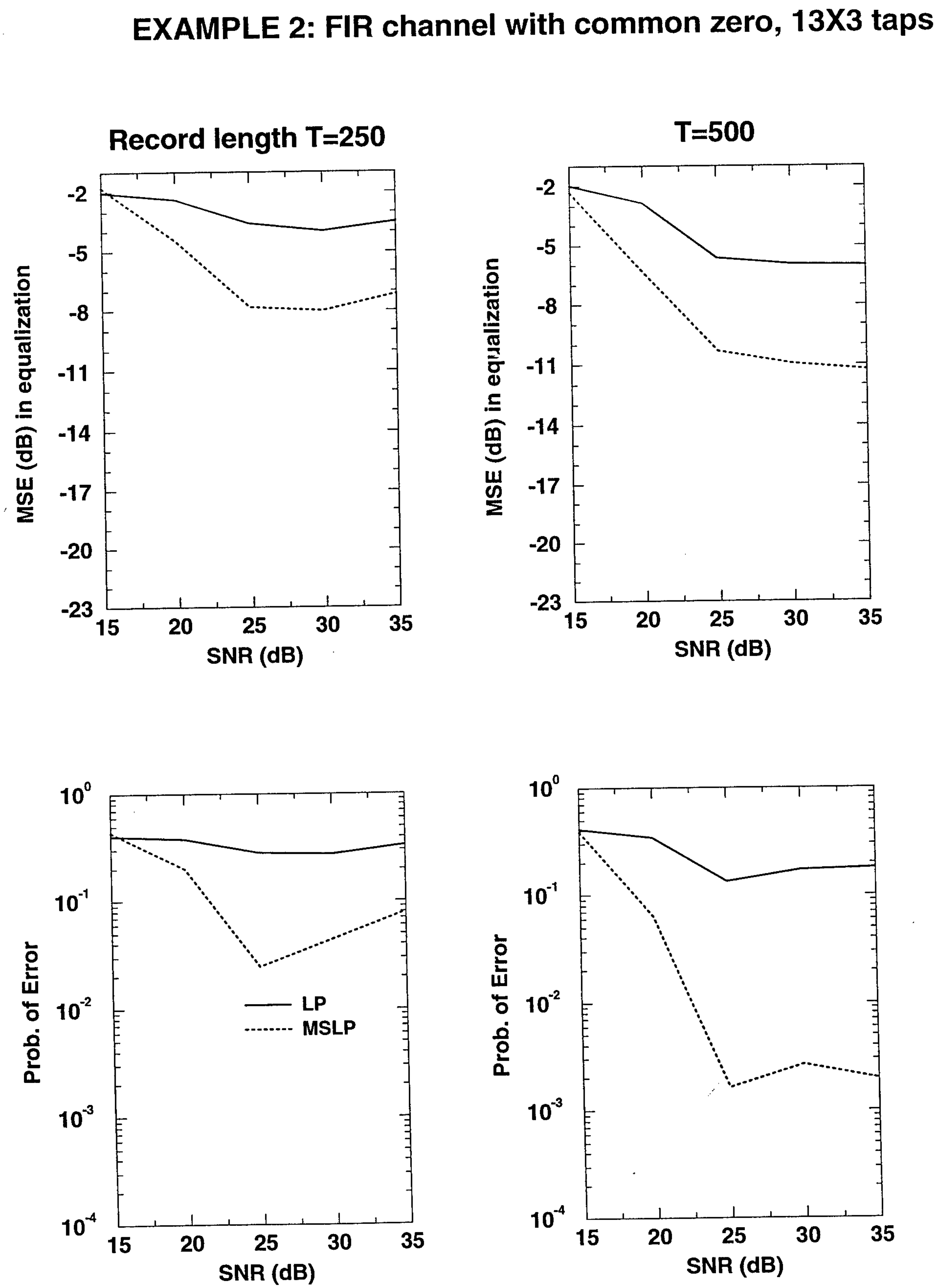


Fig. 3. Example 2: Probability of detection error and normalized MSE after equalization for various SNR's, averaged over 100 Monte Carlo runs. Left column: record length T=250 symbols; right column: record length T=500 symbols. Solid lines: Algorithm II (LP); dotted lines: Algorithm I (MSLP). Parameters: $\overline{L} = L_e = L_1 = 12$.

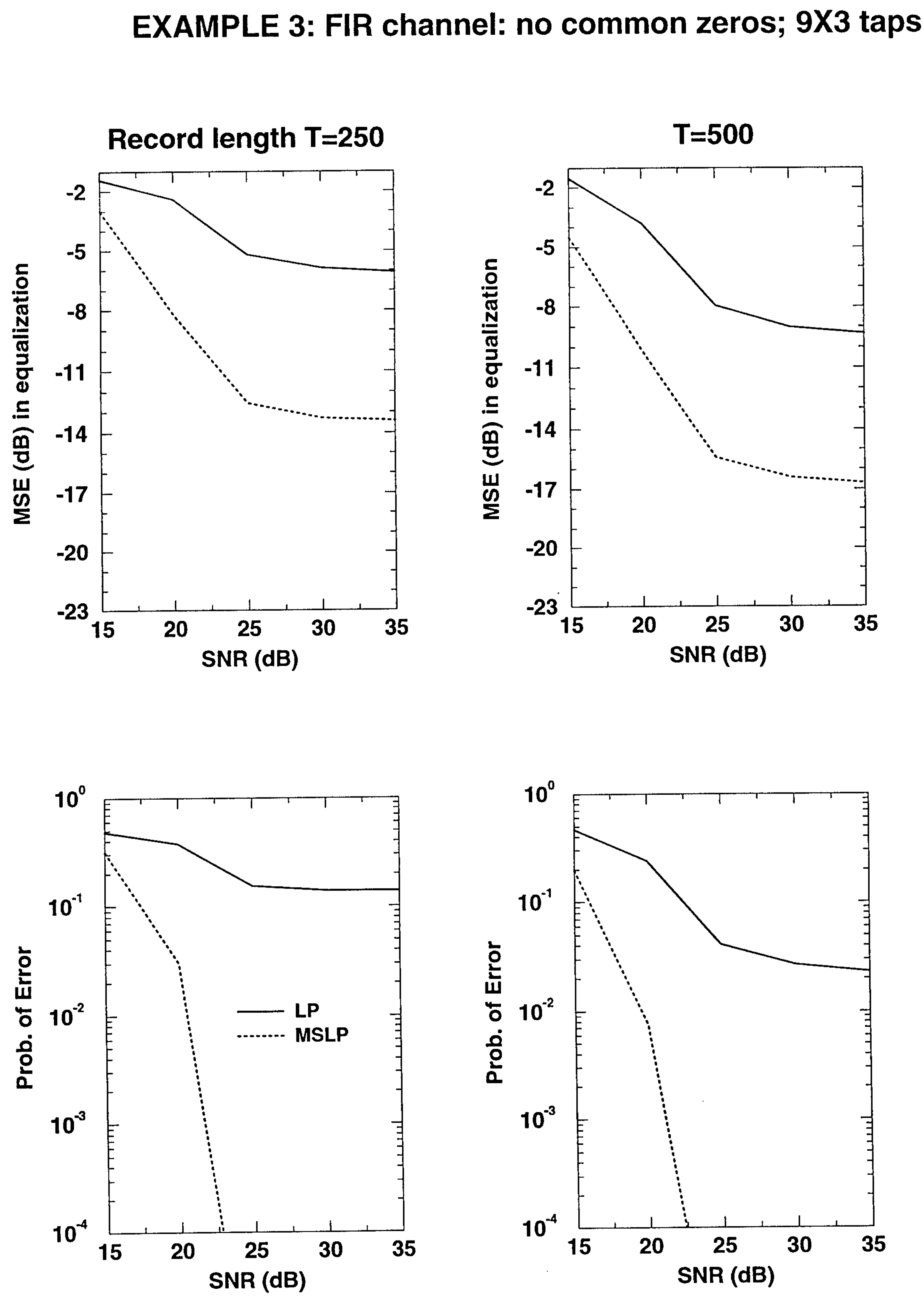


Fig. 4. Example 3: Probability of detection error and normalized MSE after equalization for various SNR's, averaged over 100 Monte Carlo runs. Left column: record length T=250 symbols; right column: record length T=500 symbols. Solid lines: Algorithm II (LP); dotted lines: Algorithm I (MSLP). Parameters: $\overline{L} = L_e = L_1 = 8$.

[5]

Adaptive Blind Separation of Convolutive Mixtures of Independent Linear Signals[1]

Jitendra K. Tugnait

Department of Electrical Engineering

Auburn University, Auburn, AL 36849, USA

Tel.: (334)844-1846 FAX: (334)844-1809

Email: tugnait@eng.auburn.edu

Abstract

This paper is concerned with the problem of blind separation of independent signals (sources) from their linear convolutive mixtures. The problem consists of recovering the sources up to shaping filters from the observations of MIMO system output. The various signals are assumed to be linear non-Gaussian but not necessarily i.i.d. (independent and identically distributed). Recently an iterative, normalized higher-order cumulant maximization based approach was developed using the fourth-order normalized cumulants of the "beamformed" data. This approach was source-iterative, i.e., the sources were extracted (at each sensor) and cancelled one-by-one, in the process yielding a decomposition of the given data at each sensor into its independent signal components. In this paper an adaptive implementation of the above approach is developed using a stochastic gradient approach. Some further enhancements including a Wiener filter implementation for signal separation and adaptive filter reinitialization are also provided. Computer simulation examples are presented to illustrate the proposed approach.

Corrected September 8, 1997.

Keywords: Spatio-temporal processing, blind signal separation, multi-input multi-output channels/systems.

[1]This work was supported by the National Science Foundation under Grant MIP-9312559 and by the Office of Naval Research under Grant N00014-97-1-0822.

1 Introduction

Given noisy measurements $y_i(k)$, $(i = 1, 2, \cdots, N)$, at time k at N sensors, let these measurements be a linear convolutive mixture of M source signals $x_j(k)$, $(j = 1, 2, \cdots, M)$:

$$y_i(k) = \sum_{j=1}^{M} U_{ij}(z)x_j(k) + n_i(k), \quad i = 1, 2, \cdots, N, \tag{1-1}$$

$$\Longrightarrow \quad \mathbf{y}(k) = \mathcal{U}(z)\mathbf{x}(k) + \mathbf{n}(k), \tag{1-2}$$

where ij-th element of $\mathcal{U}(z)$ is $U_{ij}(z)$, $\mathbf{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathbf{x}(k)$ and $\mathbf{n}(k)$, z^{-1} is both the backward-shift operator (i.e., $z^{-1}x(k) = x(k-1)$, etc.) as well as the complex variable in the $\mathcal{Z}$-transform, $x_j(k)$ is the j-th input at sampling time k, $y_i(k)$ is the i-th output, $n_i(k)$ is the additive Gaussian measurement noise, and

$$U_{ij}(z) := \sum_{l=-\infty}^{\infty} u_{ij}(l)z^{-l} \tag{1-3}$$

is the scalar transfer function with $x_j(k)$ as the input and $y_i(k)$ as the output. We allow all of the above variables to be complex-valued.

Suppose that we design a MIMO dynamic system $\mathcal{E}(z)$ with N inputs and M outputs such that the overall $M \times M$ system

$$\mathcal{T}(z) := \mathcal{E}(z)\mathcal{U}(z) \tag{1-4}$$

decouples the source signals. Following the 2×2 case considered in [7], this implies that we must have ($T_{ij}(z)$ denotes the ij-th element of $\mathcal{T}(z)$)

$$\begin{aligned} T_{ij}(z) \; &= 0 \quad \text{for} \quad i \neq i_j \\ &\neq 0 \quad \text{for} \quad i = i_j \end{aligned} \tag{1-5}$$

where $i = 1, 2, \cdots, M$; $j = 1, 2, \cdots, M$ and $i_j \in \{1, 2, \cdots, M\}$ such that $i_j \neq i_l$ for $j \neq l$. That is, in every column and every row of $\mathcal{T}(z)$ there is exactly one non-zero entry. In a blind separation problem, the nonzero entries of $\mathcal{T}(z)$ are allowed to be a scalar linear system (shaping filter), unlike the equalization problems where they must be constant gains and/or pure delays.

The problem considered above arises in a wide variety of applications: in array processing for wideband sources under multipath propagation, in speech enhancement ("cocktail party" problem), and in noise cancellation where the reference microphone does not measure noise alone ("crosstalk");

see [1]-[8], [20], [23]-[25], [29]-[33] and references therein. Blind source separation becomes necessary when the propagation between sources and sensors can not be accurately modeled for lack of knowledge of multipath environment, unknown (or imprecisely known) array manifold, etc. Separation of sources differs from blind equalization [9],[10],[13],[14],[17] in that the source signals are not necessarily i.i.d. (independent and identically distributed).

The problem of blind source separation has received increasing attention in the past few years beginning with [1]. The work done can be classified into two broad categories based upon the underlying propagation model: instantaneous mixtures and convolutive mixtures. In linear instantaneous mixture models the transfer function $\mathcal{U}(z)$ in (1-2) is a constant matrix (called the mixing matrix). This case arises for narrowband signals where any multipath produces relative delays small enough to cause just a phase shift [33]. The work reported in [1], [3]-[6], [8], [16], [29], [33] and [12] deals with this class of models; this list is by no means exhaustive, see also references therein. The general model (1-2) represents a linear convolutive mixture. The work reported in [2], [7], [17], [19], [20], [30]-[32] and [11] (and references therein) deals with linear convolutive mixture (dynamic mixing) models. In this paper we consider dynamic mixing where we allow $N \geq M$ ($N =$ number of sensors, $M=$ number of sources) with M arbitrary, whereas quite a few existing papers are restricted to either $M = N = 2$ ([7],[24],[25]) or $M = N$ ([2],[17],[11]).

Past work on separation of convolutive mixtures may be categorized into several classes: time-domain approaches ([2], [17], [19], [20], [23], [24], [25], [30], [31], [11]), frequency-domain approaches ([7],[32]), adaptive (recursive) approaches ([17], [24], [25], [30]), and non-recursive (batch) approaches ([2], [7], [8], [18]-[20], [23], [24], [30], [32], [11]). In this paper we present time-domain adaptive approaches. As noted earlier quite a few of existing approaches are limited either to $M = N = 2$ ([7], [24], [25]) or to $M = N$ ([2],[17], [11]). Although [31] and [32] treat a general case, their analyses are restricted to the case of two sources ($M = 2$). In this paper we consider a general case of $N \geq M$ with M arbitrary.

A key assumption made in this paper is that the various sources emit linear non-Gaussian signals (i.e., signals that can be represented as the output of a stable linear system driven by an i.i.d. non-Gaussian sequence); this assumption also appears in [19],[20] and [11]. This assumption is clearly satisfied by most digital communications signals. This allows one to treat the problem (or a crucial part of it) as a (blind) linear system identification problem using higher-order statistics.

Therefore, existing results on blind system identification ([9], [10], [13], [19]-[21] and [11]) become quite relevant; however, most of these papers investigate non-recursive approaches.

As noted earlier blind source separation is not blind equalization. It is separation of a dynamic mixture of signals into its independent component signals (or a linearly filtered version thereof). In [7] (and others) this is handled by assuming that (for $M = N$) the diagonal entries of $\mathcal{U}(z)$ are unity. We will not follow this approach as it is not clear how to extend this to $N > M$, and moreover, we allow a row of $\mathcal{U}(z)$ to be identically zero ("faulty" sensor) so that it may not always be possible to make such a choice. Our objective in blind separation is to decompose $\mathbf{y}(k)$ of (1-2) into its independent components $\mathcal{U}^{(i)}(z)x_i(k)$ ($\mathcal{U}^{(i)}(z)$ denotes the i–th column of $\mathcal{U}(z)$) without having a prior knowledge of $\mathcal{U}(z)$. A batch (non-adaptive) version of this paper appears in [23]. Independently of [23], a similar approach using only second-order statistics appeared in [30]. In [30] it is required that rank$\{\mathcal{U}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$) whereas we require rank$\{\mathcal{U}(z)\} = M$ only for $|z| = 1$. Note that the rank restriction of [30] implies that $N > M$ for nontrivial systems (else for $N = M$, $\mathcal{U}(z)$ will be unimodular, i.e., its determinant is a constant.) In our formulation we allow $N \geq M$. On the other hand, [30] does not require the signals $\{\mathrm{x}(k)\}$ to be non-Gaussian or linear since any stationary second-order process admits a linear representation (Wold's decomposition [15]). However, our formulation relies crucially on $\{\mathrm{x}(k)\}$ being linear non-Gaussian. Finally, given $\{\mathrm{y}(k)\}$, we also seek a minimum mean-square error (MMSE) estimate of $\mathcal{U}^{(i)}(z)x_i(k)$ whereas [30] (or [11]) does not. A system identification approach to blind source separation is followed in [19], [20] and [11]. However, the results in these papers deal with identifiability issues and no specific algorithm for source separation has been provided therein. Moreover, the MMSE source separation approach considered this paper is not considered in [19], [20] and [11].

We now turn to a brief review of past work on adaptive blind system identification as it relates to the problem under consideration. An interesting input-iterative adaptive approach using prewhitened observations and the fourth-order cumulant of the inverse-filtered data at zero-lag has been considered in [21] and [17]. The inverse filter is constrained to have a lossless filter structure which is realized using a lossless lattice filter. Such a restriction can lead to ill-conditioning of the algorithm of [21] as one iteratively extracts input sequences. A fix to this is proposed in [17] but it works only for the two-input case. Refs. [21] and [17] are restricted to 'square' systems: number of

inputs (M) equal to the number of outputs (N), whereas in this paper we allow $N \geq M$, a common occurrence in array processing. Moreover, in this paper we perform no prewhitening, rather we operate directly on the given measurements. A consequence of this is that the ill-conditioning of [21],[17] referred to above does not occur in our approach. Refs. [21] and [17] are restricted to real-valued data whereas we also consider complex-valued observations.

In [23] an iterative, inverse filter criteria based approach has been developed for blind separation of multichannel non-Gaussian processes using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach is input-iterative, i.e., the inputs are extracted and removed one-by-one. The matrix impulse response is then obtained by cross-correlating the extracted inputs with the observed outputs. A by-product of this approach is a decomposition of the given data at each sensor into its independent signal components, thereby achieving blind signal separation. In this paper we develop a stochastic gradient-based "recursification" of all of the batch optimization steps in [23].

The paper is organized as follows. The precise model assumptions and our basic approach to blind separation of convolutive mixtures are described in Sec. 2. The inverse-filter criteria-based approach of [13], the underlying identifiability results and the source separation solution implicit in the solution of [13] (see also [23]) are all briefly discussed in Sec. 3. In Sec. 4 we develop a stochastic gradient-based "recursification" of all of the batch optimization steps discussed in Sec. 3. An MMSE solution with controlled delay (d in Sec. 5), to the problem of blind signal separation given the channel impulse response estimates is discussed and analyzed in Sec. 5. Finally, two computer simulation examples are presented in Sec. 6 to illustrate the proposed approaches.

2 Model Assumptions and Signal Separation

We impose the following conditions on model (1-1)-(1-2):

(AS1) $N \geq M$, i.e., there are at least as many outputs as inputs.

(AS2) The vector sequence $\{\mathbf{x}(k)\}$ is stationary, its various components are mutually independent, and the coupling system (i.e. the transfer function $\mathcal{U}(z)$) is stable. Moreover, $\{\mathbf{x}(k)\}$ is linear, i.e.

$$\mathbf{x}(k) = \mathcal{V}(z)\mathbf{w}(k), \tag{2-1}$$

where $\{\mathrm{w}(k)\}$ is a zero-mean, $M-$vector stationary non-Gaussian process, temporally i.i.d. and spatially independent, with nonzero fourth cumulants. Because of the mutual independence of the components of $\mathrm{x}(k)$, we take $\mathcal{V}(z)$ to be diagonal.

(AS3) Consider the composite system

$$\mathbf{y}(k) = \mathcal{U}(z)\mathcal{V}(z)\mathrm{w}(k) + \mathbf{n}(k) =: \mathcal{F}(z)\mathrm{w}(k) + \mathbf{n}(k). \tag{2-2}$$

Assume that $\mathrm{rank}\{\mathcal{F}(z)\} = M$ for any $|z| = 1$.

(AS4) Since the composite system is causal, we have

$$\mathcal{F}(z) = \sum_{l=0}^{\infty} \mathrm{F}_l z^{-l} \approx \sum_{l=0}^{L} \mathrm{F}_l z^{-l}. \tag{2-3}$$

(AS5) The noise $\{\mathbf{n}(k)\}$ is a zero-mean, stationary Gaussian sequence independent of $\{\mathrm{w}(k)\}$. Moreover, it is ergodic.

Note that by (AS2) and (AS3), the statement $\mathrm{rank}\{\mathcal{F}(z)\} = M$ for any $|z| = 1$ is equivalent to the statement $\mathrm{rank}\{\mathcal{U}(z)\} = M$ for any $|z| = 1$. Since $\mathcal{F}(z)$ is stable, for IIR models (2-3) acts as a "good" approximation for large L. We will not require precise knowledge of L in the sequel. It is convenient to assume an FIR model. We will denote the $ij-$th element of $\mathcal{F}(z)$ is $\mathcal{F}_{ij}(z)$.

Let $\mathcal{F}^{(i)}(z)$ denote the $i-$th column of $\mathcal{F}(z)$. In blind convolutive signal separation we are interested in decomposing the observations at the various sensors into its independent components. That is, our objective is to estimate $\mathcal{F}^{(i)}(z)w_i(k)$ for $i = 1, 2, \cdots, M$ given $\{\mathrm{y}(k)\}$ without having a-prior knowledge of $\mathcal{F}(z)$. Note that this is different from the solutions in [2] and [7] (and others) where one obtains a "single" estimate of $\mathrm{x}_i(k)$ (or a "shaped" version of it): recall (1-4) and (1-5). By discarding all but one of the N entries of the $N-$vector $\mathcal{F}^{(i)}(z)w_i(k)$, we can get the solution specified by (1-5). Because of inherent scale and shift ambiguities (see Remark 1 in Sec. 3) we will end up estimating a scaled and shifted version of $\mathcal{F}^{(i)}(z)w_i(k)$. Thus, by assuming linearity of $\{\mathrm{x}(k)\}$ (cf. (2-1)), we have converted the blind signal separation problem into a blind MIMO channel identification and deconvolution problem to which a solution exists in [13]. It was shown in [23] as to how the solution of [13] applies to the current problem. In this paper we develop an adaptive implementation of of the approach of [23]. Also, the deconvolution solution of [13] is not necessarily an MMSE (minimum mean-square error) solution. To remedy this and to design MMSE estimators with "controlled delay" (d in Sec. 5), we also consider other modifications. Using the

channel identification results of [13], we consider designing adaptive MMSE estimates of (scaled and shifted versions) of $\mathcal{F}^{(i)}(z)w_i(k)$.

3 An Iterative Solution Based on Inverse-Filter Criteria

In [13] an iterative, inverse filter criteria based approach has been developed for deconvolution of multichannel non-Gaussian processes using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. In [23] this approach has been applied to the blind convolutive signal separation problem using a non-recursive algorithm. The approach is input-iterative, i.e., the inputs are extracted and removed one-by-one. The matrix impulse response is then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper we develop a stochastic gradient-based "recursification" of all of the batch optimization steps in [13] and [23]. In this section we briefly discuss the batch (non-recursive) approach of [23]; its adaptive version is developed in Sec. 4.

Let $\mathrm{CUM}_4(w)$ denote the fourth-order cumulant of a complex-valued scalar zero-mean random variable w, defined as

$$\mathrm{CUM}_4(w) := \mathrm{cum}_4\{w, w^*, w, w^*\} = E\{|w|^4\} - 2[E\{|w|^2\}]^2 - |E\{w^2\}|^2 \tag{3-1}$$

where $*$ denotes complex conjugation [15]. We will use the notation $\gamma_{4wi} = \mathrm{CUM}_4(w_i(k))$ and $\sigma^2_{wi} = E\{|w_i(k)|^2\}$. Consider an $1 \times N$ row-vector polynomial equalizer (filter) $\mathcal{C}^T(z)$, with its j-th entry denoted by $\mathcal{C}_j(z)$, operating on the data vector $\mathbf{y}(k)$. Let the equalizer output be denoted by $e(k)$:

$$e(k) = \sum_{i=1}^{N} \mathcal{C}_i(z) y_i(k). \tag{3-2}$$

Following [13] consider maximization of the cost

$$J := \frac{|\mathrm{CUM}_4(e(k))|}{[E\{|e(k)|^2\}]^2} \tag{3-3}$$

for designing a linear equalizer to recover one of the inputs. It is shown [13] that when (3-3) is maximized w.r.t. $\mathcal{C}(z)$, then (3-2) reduces to

$$e(k) = d w_{j_0}(k - k_0), \tag{3-4}$$

where d is some complex constant, k_0 is some integer, j_0 indexes some input out of the given M inputs, i.e., the equalizer output is a possibly scaled and shifted version of one of the system inputs. It has been established in [13] that under (AS1)-(AS4) and no noise, such a solution exists and if doubly-infinite equalizers are used, then all locally stable stationary points of the given cost w.r.t. the equalizer coefficients are also characterized by solutions such as (3-4).

An source-iterative solution is given by:

Step 1. Maximize (3-3) w.r.t. the equalizer $C(z)$ to obtain (3-4). Let

$$\gamma_{4j_0} = \mathrm{CUM}_4(e(k)) = \mathrm{CUM}_4(dw_{j_0}(k)). \tag{3-5}$$

Step 2. Cross-correlate $\{e(k)\}$ (of (3-4)) with the given data (2-2) and define a possibly scaled and shifted estimate of $f_{ij_0}(\tau)$ as

$$\widehat{f}_{ij_0}(\tau) := \frac{E\{y_i(k)e^*(k-\tau)\}}{E\{|e(k)|^2\}} \tag{3-6}$$

where $F_{ij}(z) = \sum_{l=-\infty}^{\infty} f_{ij}(l)z^{-l}$. Consider now the reconstructed contribution of $e(k)$ to the data $y_i(k)$ $(i = 1, 2, \cdots, N)$, denoted by $\widetilde{y}_{i,j_0}(k)$:

$$\widetilde{y}_{i,j_0}(k) := \sum_l \widehat{f}_{ij_0}(l)e(k-l). \tag{3-7}$$

Step 3. Remove the above contribution from the data to define the outputs of a MIMO system with N outputs and $M-1$ inputs. These are given by

$$y_i'(k) := y_i(k) - \widetilde{y}_{i,j_0}(k). \tag{3-8}$$

Step 4. If $M > 1$, set $M \leftarrow M-1$, $y_i(k) \leftarrow y_i'(k)$, and go back to **Step 1**, else quit.

In practice, all the expectations in (3-6) are replaced with their sample averages over appropriate data records.

It has been shown in [13] that

$$\widetilde{y}_{i,j_0}(k) = \sum_l f_{ij_0}(l)w_{j_0}(k-l), \tag{3-9}$$

i.e., we have decomposed the observations at the various sensors into its independent components: $\widetilde{y}_{i,j_0}(k)$ in (3-9) represents the contribution of $\{w_{j_0}(k)\}$ to the i-th sensor achieving **blind signal separation.**

Remark 1. It has been shown in [13] that under the conditions (**AS1**)-(**AS4**) and no noise, the proposed iterative approach is capable of blind identification of a MIMO transfer function

$\mathcal{F}(z)$ up to a time-shift, a scaling and a permutation matrix provided that we allow doubly-infinite equalizers. That is, given $\mathcal{F}(z)$, we end up with a $\mathcal{A}(z)$ where the two are related via

$$\mathcal{A}(z) = \mathcal{F}(z)\mathbf{DAP} \tag{3-10}$$

where $\mathbf{D}$ is an $M \times M$ "time-shift" diagonal matrix with diagonal entries such as z^{-k_0} (recall (3-4)), $\mathbf{A}$ is an $M \times M$ diagonal scaling matrix, and $\mathbf{P}$ is an $M \times M$ permutation matrix. The following result has been proved in [13].

Theorem 1[13]: Given the model (2-2) such that $\mathbf{n}(k) \equiv 0$ and given the true 4th-order and 2nd-order cumulant functions of the model output $\{\mathbf{y}(k)\}$ such that conditions **(AS1)**-**(AS4)** hold true. Suppose that doubly infinite equalizers are used in steps 1–4 of the iterative procedure of Sec. 3. Then this procedure yields a transfer function $\mathcal{A}(z)$ satisfying (3-10). • □

Remark 2. The results of [13] are based upon the use of doubly-infinite inverse filters. If we assume that $\mathcal{F}(z)$ has finite impulse response (FIR) and rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), then finite length inverse filters suffice. For an analysis and further elaborations, see [22] and [16] where a Godard cost function is considered but the results of [22] and [16] can be easily modified to apply to the cost (3-3) and the basic conclusions remain unchanged. The following result follows from [13] and [16].

Theorem 2: Given the FIR model (2-2) such that $\mathbf{n}(k) \equiv 0$ and conditions **(AS1)** and **(AS4)** hold true. Suppose that steps 1–4 of the iterative procedure of Sec. 3 are used and the record length tends to infinity. Then this procedure yields a transfer function $\mathcal{A}(z)$ satisfying (3-10) if one of the following holds true:

(A) Rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), and doubly-infinite equalizers are used.

(B) Rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), $\mathcal{F}(z)$ is column-reduced and FIR equalizers with length $L_e \geq (2M-1)L_c - 1$ are used where L_c = channel length. • □

4 Adaptive Algorithm

In this section we develop a stochastic gradient-based "recursification" of all of the batch optimization steps discussed in Sec. 3. Theorems 1 and 2 of Sec. 3 motivate and justify the algorithm

developed in this section.

4.1 First Stage Maximization of Normalized Fourth Cumulant

Let the length of the equalizer $\mathcal{C}(z)$ be L_e and let

$$\mathcal{C}_i(z) = \sum_{l=0}^{L_e-1} c_i(l) z^{-l}. \tag{4-1}$$

This allows us to rewrite (3-2) as

$$e(k) = \sum_{i=1}^{N} \sum_{l=0}^{L_e-1} c_i(l) y_i(k-l) = \mathbf{C}^T \mathbf{Y}(k) \tag{4-2}$$

where

$$\mathbf{Y}(k) = \begin{bmatrix} Y_1^T(k) & Y_2^T(k) & \cdots & Y_N^T(k) \end{bmatrix}^T, \tag{4-3}$$

$$Y_i(k) = \begin{bmatrix} y_i(k) & y_i(k-1) & \cdots & y_i(k-L_e+1) \end{bmatrix}^T, \tag{4-4}$$

$$\mathbf{C}(k) = \begin{bmatrix} C_1^T & C_2^T & \cdots & C_N^T \end{bmatrix}^T, \tag{4-5}$$

and

$$C_i = \begin{bmatrix} c_i(0) & c_i(1) & \cdots & c_i(L_e-1) \end{bmatrix}^T. \tag{4-6}$$

Define

$$m_4 = E\{|e(k)|^4\}, \quad m_2 = E\{|e(k)|^2\}, \quad \tilde{m}_2 = E\{e^2(k)\}. \tag{4-7}$$

Then showing explicit dependence upon $\mathbf{C}$, (3-3) may be rewritten as

$$J(\mathbf{C}) = \operatorname{sgn}(\gamma_4) \left[\frac{m_4 - |\tilde{m}_2|^2}{m_2^2} - 2 \right] \tag{4-8}$$

where

$$\gamma_4 = m_4 - 2\,m_2^2 - |\tilde{m}_2|^2. \tag{4-9}$$

Let $\nabla_{\mathbf{C}}$ denote a gradient operator (w.r.t. a vector $\mathbf{C}$). We will follow [26] in formally defining the complex derivatives. Then we have

$$\nabla_{\mathbf{C}^*} e(k) = 0 \quad \text{and} \quad \nabla_{\mathbf{C}^*} e^*(k) = \mathbf{Y}^*(k). \tag{4-10}$$

Using the above results in (4-7) we have

$$\nabla_{\mathbf{C}^*} m_4 \;=\; 2E\{e^2(k)e^*(k)\mathbf{Y}^*(k)\}, \quad \nabla_{\mathbf{C}^*} m_2 \;=\; E\{e(k)\mathbf{Y}^*(k)\} \tag{4-11}$$

and

$$\nabla_{\mathbf{C}^*} \tilde{m}_2 \;=\; 0, \quad \nabla_{\mathbf{C}^*} \tilde{m}_2 \;=\; 2\,E\{e^*(k)\mathbf{Y}^*(k)\}. \tag{4-12}$$

Using (4-8)-(4-12) and after some simplification, we have

$$\nabla_{\mathbf{C}^*} J(\mathbf{C}) \;=\;$$

$$\frac{2\,\mathrm{sgn}(\gamma_4)}{m_2^3}\left\{m_2 E\{|e(k)|^2 e(k)\mathbf{Y}^*(k)\} - \tilde{m}_2 m_2 E\{e^*(k)\mathbf{Y}^*(k)\} - [m_4 - |\tilde{m}_2|^2]E\{e(k)\mathbf{Y}^*(k)\}\right\}. \tag{4-13}$$

We will use a stochastic gradient method for recursification of maximization of $J(\mathbf{C})$ using an 'instantaneous' gradient as an estimate of (4-13). Given the estimate $\mathbf{C}(k-1)$ of the tap-gains at time $k-1$, the stochastic gradient method computes the update $\mathbf{C}(k)$ at time k as

$$\tilde{\mathbf{C}}(k) \;=\; \mathbf{C}(k-1) + \mu_1 \,\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k-1)) \tag{4-14}$$

$$\mathbf{C}(k) \;=\; \frac{\tilde{\mathbf{C}}(k)}{||\tilde{\mathbf{C}}(k)||} \tag{4-15}$$

where μ_1 is the update step-size and $\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k-1))$ is an instantaneous gradient of the cost J (w.r.t. $\mathbf{C}^*$) at time k evaluated at $\mathbf{C}(k-1)$. Since the cost J is invariant any scaling of $\mathbf{C}$, we normalize $\mathbf{C}$ in (4-15) to have a unit norm. From (4-13) we have the approximation

$$\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k)) \;=\; \mathrm{sgn}(\gamma_{4k})\frac{2}{m_{2k}^3}\left\{\left[m_{2k}\left(e^2(k) - \tilde{m}_{2k}\right)e^*(k) - \left(m_{4k} - |\tilde{m}_{2k}|^2\right)e(k)\right]\mathbf{Y}^*(k)\right\} \tag{4-16}$$

where

$$m_{2k} \;=\; (1-\mu_2)m_{2(k-1)} + \mu_2|e(k)|^2, \tag{4-17}$$

$$\tilde{m}_{2k} \;=\; (1-\mu_2)\tilde{m}_{2(k-1)} + \mu_2 e^2(k), \tag{4-18}$$

$$m_{4k} \;=\; (1-\mu_2)m_{4(k-1)} + \mu_2|e(k)|^4, \tag{4-19}$$

$$\gamma_{4k} \;=\; m_{4k} - 2\,m_{2k}^2 - |\tilde{m}_{2k}|^2 \tag{4-20}$$

and

$$e(k) \;=\; \mathbf{C}^T(k)\mathbf{Y}(k). \tag{4-21}$$

In (4-17)-(4-19) the various quantities represent estimates based upon sample averaging, the (exponential window) memory being controlled by the forgetting factor μ_2 ($0 < \mu_2 < 1$). The initializations for (4-17)-(4-19) are: $m_{20} = m_{40} = \tilde{m}_{20} = 0$.

4.2 First Stage Signal Cancellation

Now we discuss implementation of (3-6) via sample averaging using an exponential window controlled by a forgetting factor μ_3. Define $(L_1, L_2 > 0)$

$$\{\tilde{f}_i(n)\}_{n=-L_1}^{L_2} \;=\; \text{impulse response from } \{e(k)\} \text{ (input) to } \{y_i(k)\} \text{ (output)} \tag{4-22}$$

and

$$\widetilde{\mathbf{F}}_n = \left[\; \tilde{f}_1(n) \;\; \tilde{f}_2(n) \;\; \cdots \;\; \tilde{f}_N(n) \;\right]^T . \tag{4-23}$$

By Sec. 3, when (4-8) is maximized, $e(k)$ satisfies (3-4) so that for suitable choice of L_1 and L_2, there exists a $j_0 \in \{1, 2, \cdots, M\}$ such that

$$\sum_l f_{ij_0}(l) w_{j_0}(k-l) \;=\; \left[\sum_{n=-L_1}^{L_2} \widetilde{\mathbf{F}}_n^T e(k-n) \right]_{i1} , \quad i = 1, 2, \cdots, N, \tag{4-24}$$

where $[\mathbf{A}]_{ij}$ denotes the $ij-$th element of the matrix $\mathbf{A}$. Note that (4-24) (cf. (3-7)) represents the contribution of the extracted source at stage 1 to the measurement at time k at the $i-$th sensor. In order to implement (3-7) and (3-8), we need recursive estimates of $\widetilde{\mathbf{F}}_n$. The estimate $\widetilde{\mathbf{F}}_n(k)$ of $\widetilde{\mathbf{F}}_n$ at time k is provided by

$$\widetilde{\mathbf{F}}_n(k) \;=\; \mathbf{R}_n(k)/m_{ee}(k) \tag{4-25}$$

where

$$m_{ee}(k) \;=\; (1-\mu_3)m_{ee}(k-1) \;+\; \mu_3|e(k)|^2, \tag{4-26}$$

$$\mathbf{R}_n(k) \;=\; (1-\mu_3)\mathbf{R}_n(k-1) \;+\; \mu_3 \mathbf{y}(k)e^*(k-n). \tag{4-27}$$

4.3 Multistage Algorithm

In Secs. 4.1 and 4.2 we discussed the first stage of the algorithm where we have N sensors and M sources. Now we put it all together following the source-iterative solution of Sec. 3 and discuss extraction of M sources including the cancellation of the extracted sources. We will use the superscript (m) to denote the various quantities pertaining to stage m. These have been used previously in Secs. 4.1 and 4.2 without this superscript; for instance, $\mathbf{C}^{(m)}(k)$ now denotes the estimate of the tap-gain vector at time k at stage m, etc.

Initialization:

$$\mathbf{Y}^{(1)}(k) \;=\; \text{as in (4-3), and } \mathbf{y}^{(1)}(k) \;=\; \mathbf{y}(k). \tag{4-28}$$

DO FOR $m = 1, 2, \cdots, M$:

$$\widetilde{\mathbf{C}}^{(m)}(k) \;=\; \mathbf{C}^{(m)}(k-1) + \mu_1 \, \nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k-1)) \tag{4-29}$$

$$\mathbf{C}^{(m)}(k) \;=\; \frac{\widetilde{\mathbf{C}}^{(m)}(k)}{\|\widetilde{\mathbf{C}}^{(m)}(k)\|} \tag{4-30}$$

where

$$\nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k)) \;=\; \operatorname{sgn}(\gamma_{4k}^{(m)}) \frac{2}{m_{2k}^{(m)3}} \left\{ \left[m_{2k}^{(m)} \left(e^{(m)2}(k) - \tilde{m}_{2k}^{(m)} \right) e^{(m)*}(k) \right.\right.$$

$$\left.\left. - \left(m_{4k}^{(m)} - |\tilde{m}_{2k}^{(m)}|^2 \right) e^{(m)}(k) \right] \mathbf{Y}^{(m)*}(k) \right\}, \tag{4-31}$$

$$m_{2k}^{(m)} \;=\; (1-\mu_2) m_{2(k-1)}^{(m)} + \mu_2 |e^{(m)}(k)|^2, \tag{4-32}$$

$$\tilde{m}_{2k}^{(m)} \;=\; (1-\mu_2) \tilde{m}_{2(k-1)}^{(m)} + \mu_2 e^{(m)2}(k), \tag{4-33}$$

$$m_{4k}^{(m)} \;=\; (1-\mu_2) m_{4(k-1)}^{(m)} + \mu_2 |e^{(m)}(k)|^4, \tag{4-34}$$

$$\gamma_{4k}^{(m)} \;=\; m_{4k}^{(m)} - 2\, m_{2k}^{(m)2} - |\tilde{m}_{2k}^{(m)}|^2 \tag{4-35}$$

and

$$e^{(m)}(k) \;=\; \mathbf{C}^{(m)T}(k) \mathbf{Y}^{(m)}(k). \tag{4-36}$$

Set

$$\widehat{\tilde{\mathbf{y}}}^{(m)}(k) \;=\; \sum_{n=-L_1}^{L_2} \widetilde{\mathbf{F}}_n^{(m)}(k) e^{(m)}(k-n) \tag{4-37}$$

where $\widehat{\tilde{\mathbf{y}}}^{(m)}(k)$ represents (cf. (3-7)) the contribution of the extracted source at the m−th stage to the measurements at time k, and where $(n = -L_1, -L_1+1, \cdots, L_2)$

$$\widetilde{\mathbf{F}}_n^{(m)}(k) \;=\; \mathbf{R}_n^{(m)}(k) / m_{ee}^{(m)}(k), \tag{4-38}$$

$$m_{ee}^{(m)}(k) \;=\; (1-\mu_3) m_{ee}^{(m)}(k-1) + \mu_3 |e^{(m)}(k)|^2, \tag{4-39}$$

$$\mathbf{R}_n^{(m)}(k) \;=\; (1-\mu_3)\mathbf{R}_n^{(m)}(k-1) \,+\, \mu_3 \mathbf{y}^{(m)}(k) e^{(m)*}(k-n) \tag{4-40}$$

and

$$\mathbf{y}^{(m+1)}(k) \;=\; \mathbf{y}^{(m)}(k) \,-\, \widehat{\widetilde{\mathbf{y}}}^{(m)}(k). \tag{4-41}$$

Define

$$\widetilde{Y}_i^{(m)}(k) = \left[\; \widetilde{y}_i^{(m)}(k) \quad \widetilde{y}_i^{(m)}(k-1) \quad \cdots \quad \widetilde{y}_i^{(m)}(k-L_e+1) \; \right]^T \tag{4-42}$$

where $\widetilde{y}_i^{(m)}(k)$ denotes the $i-$th component of $\widetilde{\mathbf{y}}^{(m)}(k)$. Set

$$\mathbf{Y}^{(m+1)}(k) = \left[\; Y_1^{(m+1)T}(k) \quad Y_2^{(m+1)T}(k) \quad \cdots \quad Y_N^{(m+1)T}(k) \; \right]^T \tag{4-43}$$

where

$$Y_i^{(m+1)}(k) \;=\; Y_i^{(m)}(k) \,-\, \widetilde{Y}_i^{(m)}(k). \tag{4-44}$$

ENDDO

The sequence $\{\widehat{\widetilde{\mathbf{y}}}^{(m)}(k)\}$ in (4-37) represents the contribution of the extracted source at the $m-$th stage to the measurements at time k.

Remark 3. If M were unknown the proposed approach will still work in the sense that if M were underestimated, some sources will be missed but the extracted sources will correspond to one of the sources. If M were overestimated, all the sources will be recovered in addition to some "meaningless junk" outputs in stages M_0+1 and later where M_0 denotes true number of sources. Indeed one can test the 'residuals' (4-44) (see also (3-8)) to check if any significant non-Gaussian components remain in the data before implementing another equalizer in parallel. We do not pursue this aspect in this paper. □

Running Cost. To monitor the convergence of the equalizers in various stages of the algorithm, it is useful to calculate a running cost (4-8) without the sign. Let $J_k^{(m)}$ denote the running cost for the $m-$th stage at time k, given by

$$J_k^{(m)} \;=\; \frac{m_{4k}^{(m)} - |\widetilde{m}_{2k}^{(m)}|^2}{m_{2k}^{(m)2}} - 2 \tag{4-45}$$

where

$$m_{2k}^{(m)} = (1-\mu_4)m_{2(k-1)}^{(m)} + \mu_4|e^{(m)}(k)|^2, \tag{4-46}$$

$$\tilde{m}_{2k}^{(m)} = (1-\mu_4)\tilde{m}_{2(k-1)}^{(m)} + \mu_4 e^{(m)2}(k), \tag{4-47}$$

and

$$m_{4k}^{(m)} = (1-\mu_4)m_{4(k-1)}^{(m)} + \mu_4|e^{(m)}(k)|^4. \tag{4-48}$$

For all of the simulations presented in Sec. 6, we took $\mu_4 = 0.002$.

5 Further Modifications

5.1 MMSE Signal Separation

5.1.1 Non-recursive Processing

Recall that our objective is to estimate $\mathcal{F}^{(i)}(z)w_i(k)$ for $i = 1, 2, \cdots, M$ given $\{\mathrm{y}(k)\}$. The non-recursive solution of Sec. 3 provides a solution in the form of (3-7) (see also (3-9)) whereas the adaptive solution of Sec. 4 has it as (4-37). The deconvolution solution of [13] is not necessarily an MMSE solution. It has been shown in [35] and [36] (and references therein) that for the constant modulus algorithm, under certain conditions, the resultant solution (extracted source in the first stage) may be "close" to an MMSE solution. It is possible that a similar result may hold for the problem under consideration here. However, even if it were true, the resulting solution may not be the best possible because the performance of an MMSE solution depends upon the "delay" (d in the sequel) used and the blind algorithms provide no control over the choice of the delay parameter [36]. A by-product of the solutions of Secs. 3 and 4 is the estimates of the system/channel impulse response. These estimates can be used to design MMSE estimators of $\mathcal{F}^{(i)}(z)w_i(k)$ with a controlled delay d to obtain an "optimum" performance. These considerations are nevertheless heuristic as we are ignoring any effects of additive noise on the channel estimates.

Let $\mathbf{F}_l^{(i)}$ denote the i-th column of $\mathbf{F}_l$. We wish to design a linear MMSE filter (equalizer) of length $L_e + 1$ to estimate $\tilde{\mathrm{y}}^{(j)}(k-d)$ as $\hat{\tilde{\mathrm{y}}}^{(j)}(k-d)$ given $\mathrm{y}(l)$ for $l = k, k-1, \cdots, k-L_e+1$ where $d \geq 0$,

$$\tilde{\mathrm{y}}^{(j)}(k) := \mathcal{F}^{(j)}(z)w_j(k) = \sum_{l=0}^{L} \mathbf{F}_l^{(j)} w_j(k-l) \tag{5-1}$$

and

$$\widehat{\widetilde{\mathbf{y}}}^{(j)}(k-d) \;:=\; \sum_{i=0}^{L_e-1} \mathbf{G}_i \mathbf{y}(k-i). \tag{5-2}$$

Both L_e and the delay d are "pre-determined." Using the orthogonality principle [27], the normal equations for MMSE estimator are given by

$$E\left\{[\widehat{\widetilde{\mathbf{y}}}^{(j)}(k-d) - \widetilde{\mathbf{y}}^{(j)}(k-d)]\mathbf{y}^{\mathcal{H}}(l)\right\} \;=\; 0 \tag{5-3}$$

for $l = k - L_e + 1, k - L_e + 2, \cdots, k$ where $\mathcal{H}$ denotes the Hermitian operation (complex conjugate transpose). Using (2-2), (2-3), (5-1), (5-2) and assumption (AS2), and assuming that the system model is completely known, after some manipulations (5-3) simplifies to

$$\sum_{i=0}^{L_e-1} \mathbf{G}_i \mathbf{R}_{yy}(p-i) \;=\; \sigma_{wj}^2 \sum_{k=0}^{L} \mathbf{F}_k^{(j)} \mathbf{F}_{k+d-p}^{(j)\mathcal{H}} \;=\; \sigma_{wj}^2 \mathbf{H}_{d-p}, \quad p = 0, 1, \cdots, L_e - 1 \tag{5-4}$$

where

$$\mathbf{H}_{d-p} \;:=\; \sum_{k=0}^{L} \mathbf{F}_k^{(j)} \mathbf{F}_{k+d-p}^{(j)\mathcal{H}} \;=\; \sum_{k=0}^{L} \mathbf{F}_{k+p-d}^{(j)} \mathbf{F}_k^{(j)\mathcal{H}} \;=\; \sigma_{wj}^{-2} E\left\{\widetilde{\mathbf{y}}^{(j)}(k+p-d)\widetilde{\mathbf{y}}^{(j)\mathcal{H}}(k)\right\} \tag{5-5}$$

and

$$\mathbf{R}_{yy}(p) \;:=\; E\{\mathbf{y}(t+p)\mathbf{y}^{\mathcal{H}}(t)\}. \tag{5-6}$$

Note that a shift in the sequence $\{\mathbf{F}_k^{(j)}\}$ leaves $\mathbf{H}_{d-p}$ unaffected. The desired solution when the model is completely known is therefore given by

$$\begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \cdots & \mathbf{G}_{L_e-1} \end{bmatrix} = \sigma_{wj}^2 \begin{bmatrix} \mathbf{H}_d & \mathbf{H}_{d-1} & \cdots & \mathbf{H}_{d-L_e} \end{bmatrix} \mathcal{R}_{yy}^{-1} \tag{5-7}$$

where $\mathcal{R}_{yy}$ denotes a $[NL_e] \times [NL_e]$ correlation matrix with $\mathbf{R}_{yy}(j-i)$ as its ij-th block element.

In order to obtain a data-based solution, we simply replace all the unknowns by their estimates. Since there is an inherent scale ambiguity in estimating the composite channel impulse response (cf. (3-10)), we design the equalizer only up to a scale factor by omitting σ_{wj}^2 from (5-7).

Remark 4. *Selection of Delay d:* In designing (5-2) the delay d was pre-determined. It is well-known [28] that the choice of d has a strong influence on the resultant mean-square error. One may choose to select d via exhaustive optimization as detailed below. Using the orthogonality principle [27], the MMSE when (5-2) is used is given by

$$\mathcal{J}(d) \;:=\; -\mathrm{tr}E\left\{[\widehat{\widetilde{\mathbf{y}}}^{(j)}(k-d) - \widetilde{\mathbf{y}}^{(j)}(k-d)]\widetilde{\mathbf{y}}^{(j)\mathcal{H}}(k-d)\right\} \tag{5-8}$$

where tr stands from trace. Using (5-1)-(5-7), we can simplify (5-8) to

$$\mathcal{J}(d) := \operatorname{tr} E\left\{\tilde{\mathbf{y}}^{(j)}(k-d)\tilde{\mathbf{y}}^{(j)\mathcal{H}}(k-d)\right\} - \mathcal{J}'(d) \tag{5-9}$$

where

$$\mathcal{J}'(d) := \sigma_{wj}^4 \operatorname{tr} \mathcal{H}\mathcal{R}_{yy}^{-1}\mathcal{H}^{\mathcal{H}} \tag{5-10}$$

and

$$\mathcal{H} := \begin{bmatrix} \mathbf{H}_d & \mathbf{H}_{d-1} & \cdots & \mathbf{H}_{d-L_e} \end{bmatrix}. \tag{5-11}$$

Since the first term on the right-side of (5-9) is independent of d, minimizing $\mathcal{J}(d)$ w.r.t. d is equivalent to maximizing $\mathcal{J}'(d)$ or $\sigma_{wj}^{-4}\mathcal{J}'(d)$. In practice, we replace the unknowns in (5-10) with their estimates. □

5.1.2 Adaptive Implementation

We now turn to an adaptive implementation of (5-7). Note that $\mathcal{R}_{yy}^{-1}$ does not depend upon the stage m of the algorithm of Sec. 4.3; it depends solely upon the measured data. Its computation can easily be recursified by using the matrix inversion lemma: see Table 13.1 on p. 569 in [34]; we omit the details. Denote the data-based adaptive estimate of $\mathcal{R}_{yy}^{-1}$ at time k as $\mathcal{P}_{yy}(k)$.

Let $\mathbf{H}_l^{(m)}(k)$ denote the estimate of $\mathbf{H}_l$ at stage m and time k of the multistage algorithm of Sec. 4.3. Note that $\tilde{\mathbf{F}}_n^{(m)}(k)$ in (4-38) (see also (3-6), (4-23) and (4-25)) denotes an estimate of $\mathbf{F}_n^{(i)}$ for some $i \in \{1, 2, \cdots, M\}$ (up to a scale factor and time shift, cf. Theorem 1). Therefore, from (5-5) we have the adaptive implementation at stage m as

$$\mathbf{H}_l^{(m)}(k) = \sum_{n=-L_1}^{L_2} \tilde{\mathbf{F}}_n^{(m)}(k)\tilde{\mathbf{F}}_{n+l}^{(m)\mathcal{H}}(k), \quad l = d, d-1, \cdots, d-L_e+1. \tag{5-12}$$

Combining the above two results, the adaptive MMSE estimate with lag d, $\hat{\tilde{\mathbf{y}}}^{(m)}(k)$, at stage m (corresponding to $\{\tilde{\mathbf{y}}^{(m)}(k)\}$ in (4-37)) is given by

$$\hat{\tilde{\mathbf{y}}}^{(m)}(k) := \sum_{i=0}^{L_e-1} \mathbf{G}_i^{(m)}(k)\mathbf{y}(k-i) \tag{5-13}$$

where

$$\begin{bmatrix} \mathbf{G}_0^{(m)}(k) & \mathbf{G}_1^{(m)}(k) & \cdots & \mathbf{G}_{L_e-1}^{(m)}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_d^{(m)}(k) & \mathbf{H}_{d-1}^{(m)}(k) & \cdots & \mathbf{H}_{d-L_e+1}^{(m)}(k) \end{bmatrix} \mathcal{P}_{yy}(k). \tag{5-14}$$

Eqn. (5-13) provides an approximately MMSE blind signal separation solution at stage m and time k. [Note that we have ignored any effects of additive noise on the channel estimates.]

Selection of "optimum" d as discussed in Remark 4 recursifies in an obvious way; therefore, the details are omitted.

5.2 Adaptive Filter Reinitialization

In the source-iterative (multistage) approaches of Secs. 3 and 4, any errors in cancelling the extracted sources from the preceding stages $l = 1, 2, \cdots, m-1$ affect the performance at stage m. The only stage that is immune to this phenomenon is stage $m = 1$. The multistage approach is used to make sure that each stage converges to a distinct source. A possible solution to alleviate this error propagation from stage-to-stage is to use parallel stages where we still have M stages for M sources but they all operate directly on the given data record in parallel but with different initializations of the equalizers (filters). The basic problem with such an approach is how to ensure that each stage converges to a distinct source. Here we propose to initialize the parallel stages using the results of the serial multistage implementation of Sec. 4.3 coupled with an MMSE solution similar to that of Sec. 5.1. A similar though not identical approach has been proposed in [35] in a slightly different context where the MMSE initializer has not been used.

For stage $m = 1$, there are no changes to the algorithm of Sec. 4.3. For stages $m \geq 2$, run the algorithm of Sec. 4.3 till the running cost (4-45) reaches a steady-state. Given the estimates of the subchannel impulse response at stage m, we can design an MMSE filter (in a fashion similar to Sec. 5.1.2) to estimate $w_j(k-d)$ given $\mathbf{y}(l)$ for $l = k, k-1, \cdots, k-L_e+1$. Let the extracted $w_j(k)$ at stage m be denoted by $w^{(m)}(k)$. Mimicking Sec. 5.1.2, a recursive MMSE solution at stage m and time k is given by

$$\widehat{w}^{(m)}(k-d) := \sum_{i=0}^{L_e-1} \overline{\mathbf{G}}_i^{(m)}(k)\mathbf{y}(k-i) \tag{5-15}$$

where

$$\begin{aligned} &\left[\ \overline{\mathbf{G}}_0^{(m)}(k) \quad \overline{\mathbf{G}}_1^{(m)}(k) \quad \cdots \quad \overline{\mathbf{G}}_{L_e-1}^{(m)}(k)\ \right] \\ &\qquad = \left[\ \widetilde{\mathbf{F}}_d^{(m)\mathcal{H}}(k) \quad \widetilde{\mathbf{F}}_{d-1}^{(m)\mathcal{H}}(k) \quad \cdots \quad \widetilde{\mathbf{F}}_0^{(m)\mathcal{H}}(k) \quad 0 \quad \cdots \quad 0\ \right] \mathcal{P}_{yy}(k). \end{aligned} \tag{5-16}$$

At stage m and time k, $\widehat{w}^{(m)}(k-d)$ is an MMSE estimate (with delay d) of $e^{(m)}(k)$ for the parallel implementation. Comparing (4-2) with (5-14) we see that $\mathcal{C}(z) = \sum_{i=0}^{L_e-1} \overline{\mathbf{G}}_i^{(m)}(k)z^{-i}$ is the desired MMSE initializer.

Selection of "optimum" d mimicking Remark 4 recursifies in an obvious way; therefore, the details are omitted.

6 Simulation Example

In this section we present two simulation examples to illustrate the proposed approaches. In both the examples $\mathbf{F}_0$ is of rank $1 < M = 2$ implying that rank$\{\mathcal{F}(z)\} < M = 2$ so that $\mathcal{F}(z)$ does not have a finite-length left inverse. In both the examples the length of the inverse filters was 11 samples per sensor (output) for the approach of Sec. 4. The proposed approach was applied with $M = 2$ inverse filters and $M - 1 = 1$ signal cancellers running in parallel, each successive inverse filter put in operation after waiting for 200 samples w.r.t. the previous stage. To design a channel estimate-based MMSE signal separator (following Sec. 5), we chose the length of the MMSE filter as $L_e + 1 = 11$, the same as for the inverse filters of Sec. 4. Furthermore, we chose the delay d for the MMSE separator design by following the procedure outlined in Remark 4 in Sec. 5.

6.1 Example 1

Consider a 2-input 2-output MA(6) system model resulting in N=2 and M=2 in (2-2). Its 2×2 transfer function $\mathcal{F}(z)$ was chosen as

$$\begin{bmatrix} 0.2 + 0.8z^{-1} + 0.4z^{-2} & 0.5 - 0.3z^{-1} \\ 0.3z^{-1} - 0.6z^{-2} & -0.21z^{-1} - 0.5z^{-2} + 0.72z^{-3} + 0.36z^{-4} + 0.21z^{-6} \end{bmatrix}. \tag{6-1}$$

The input $\{w_1(k)\}$ is an i.i.d. complex Gaussian-mixture (independent and identically distributed real and imaginary parts with the real part being $\mathcal{N}(0,1)$ with probability 0.9 and $\mathcal{N}(0,4)$ with probability 0.1) with 4th normalized cumulant as 0.7433 . The input $\{w_2(k)\}$ is an i.i.d. 4-QAM sequence with 4th normalized cumulant as -1. The additive noise is temporally and spatially white, zero-mean, complex Gaussian distributed (independent real and imaginary parts). The powers of

$\{w_j(k)\}$ for $j = 1$ and 2 were scaled so as to have $E\{\|\mathcal{F}^{(1)}(z)w_1(k)\|^2\} = E\{\|\mathcal{F}^{(2)}(z)w_2(k)\|^2\}$.

For signal separation the performance measure was taken to be the signal-to-interference-and-noise ratio (SINR) per source signal, defined as

$$\mathrm{SINR}_j = \frac{E\{\|\tilde{\mathbf{y}}^{(j)}(k)\|^2}{E\{\|\tilde{\mathbf{y}}^{(j)}(k) - \hat{\alpha}\hat{\tilde{\mathbf{y}}}^{(j)}(k)\|^2\}} \tag{6-2}$$

where $\hat{\alpha}$ is that value of the scalar α which minimizes $E\{\|\tilde{\mathbf{y}}^{(j)}(k) - \alpha\hat{\tilde{\mathbf{y}}}^{(j)}(k)\|^2\}$; this is need to remove the scale ambiguity in the design of (5-4) – it doesn't affect the SINR. As noted in Sec. 5, that the shift ambiguities in estimating $\mathbf{F}_k^{(j)}$ do not have any influence on the equalizer design, hence on (6-2).

The adaptive approach of Sec. 4.3 was applied without as well as with the reinitialization of Sec. 5.2. The average signal-to-noise ratio (SNR) per source was taken to be 27 dB, 20 dB, 14 dB and 7 dB, respectively, in four sets of 100 Monte Carlo runs where the SNR for a given source $\mathcal{F}^{(j)}(z)w_j(k)$ is defined as

$$\mathrm{SNR} = \frac{\frac{1}{N}E\{\|\mathcal{F}^{j}(z)w_j(k)\|^2\}}{E\{|n_i(k)|^2\}}. \tag{6-3}$$

The initial guess for the tap gains was taken to be center-tap initialization: set $c_i(5) = 1$ for $i = m$ for the $m-$th stage equalizer ($m = 1, 2$) with the remaining tap gains set to zero. The algorithm step sizes and forgetting factors for each stage m were chosen as: $\mu_1 = 0.0005$ in (4-29), $\mu_2 = 0.015$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41) when $\gamma_{4k}^{(m)} \leq 0$ (see (4-35)), and $\mu_1 = 0.00025$ in (4-29), $\mu_2 = 0.0075$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41) when $\gamma_{4k}^{(m)} > 0$. For the running cost (4-45) computation we selected $\mu_3 = 0.002$ in (4-46)-(4-48). The parameters L_1 and L_2 in (4-37) (see also (4-22) and (4-23)) were selected as $L_1 = 15$ and $L_2 = 6$. To design the MMSE equalizers/filters (5-13) or (5-15) we took $L_e = 11$ and d was optimized following Remark 4 of Sec. 5.1.1 over the range $[-15, 6]$.

Fig. 1 shows the evolution of the average running cost $J_k^{(m)}$ (see (4-45)), averaged over 100 Monte Carlo runs (after 'assigning' each equalizer cost to its corresponding extracted source) without using any filter reinitialization. For the 4-QAM sources the 4th-order normalized cumulant equals -1; therefore, at convergence, the running cost (4-45) should be close to -1. In Fig. 1 we see these values to be less than that which is largely a consequence of noise in the data which affects only the denominator of (4-45) making it larger than it should be. Similar effect is seen for the Gaussian

mixture source whose 4th-order normalized cumulant equals 0.7433 . Fig. 2 shows the evolution of the average running cost $J_k^{(m)}$ when reinitialization (after 12000 samples) of Sec. 5.2 is used. It turns out that source 1 ($w_1(k)$: Gaussian mixture) is extracted first, so that reinitialization only affects source 2 (4-QAM).

Table I shows the average SINR (averaged over 100 Monte Carlo runs) for the two sources (as per (6-2)) at the end of the run (i.e. at $k = 18000$) without and with filter reinitialization, for various SNR's. The SINR's were computed using the solution (4-37) as well as the MMSE solution of Sec. 5.1.2. It is seen that blind signal separation benefits from both, MMSE signal separation as well as filter reinitialization.

6.2 Example 2

Consider a 2-input 3-output MA(6) system model resulting in N=3 and M=2 in (2-2). Its 3×2 transfer function $\mathcal{F}(z)$ was chosen as

$$\begin{bmatrix} 0.2+0.8z^{-1}+0.4z^{-2} & 0.5-0.3z^{-1} \\ 0.3z^{-1}-0.6z^{-2} & -0.21z^{-1}-0.5z^{-2}+0.72z^{-3}+0.36z^{-4}+0.21z^{-6} \\ 0. & 0. \end{bmatrix}. \tag{6-4}$$

Notice that the last row of (6-1) is identically zero signifying that the third 'sensor' is not receiving any information signal, just noise. The first two rows of (6-4) are identical to (6-1). The inputs $\{w_j(k)\}$ ($j = 1, 2$) and additive noise are as in Example 1. The powers of $\{w_j(k)\}$ for $j = 1$ and 2 were scaled as in Example 1 to achieve equal average signal power at the sensors. The measurement SNR's defined as in (6-3) and they were selected as 25.2 dB, 18.2 dB, 12.2 dB and 5.2 dB, respectively, in four sets of 100 Monte Carlo runs.

The adaptive approach of Sec. 4.3 was applied without as well as with the reinitialization of Sec. 5.2. The various parameters chosen for signal separation were exactly as for Example 1. Figs. 3 and 4 are the counterparts to Figs. 1 and 2, respectively, of Example 1, and Table II is the counterpart to Table I of Example 1. As in Example 1, it is seen that blind signal separation benefits from both, MMSE signal separation as well as filter reinitialization. Comparing with the results of Example 1, it is seen that the results of Examples 1 and 2 are quite close to each other inspite of having a third "misleading" sensor in Example 2 that measures just noise.

7 Conclusions

The problem of blind separation of independent linear non-Gaussian signals (sources) from their linear convolutive mixtures was considered. In [23] an iterative, normalized higher-order cumulant maximization based approach was developed using the third-order and/or fourth-order normalized cumulants of the "beamformed" data. The approach is source-iterative, i.e., the sources are extracted (at each sensor) and cancelled one-by-one, providing a decomposition of the given data at each sensor into its independent signal components. In this paper we developed a stochastic gradient-based recursification of all of the batch optimization steps in [23].

Some further modifications and enhancements were also considered. For blind signal separation the estimated channel was used to decompose the received signal at each sensor into its independent signal components via an MMSE filter with a controlled delay. The proposed blind adaptive algorithm and its variations were illustrated via two simulation examples.

8 References

[1] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromorphic architecture," *Signal Processing*, vol. 24, pp. 1-10, 1991.

[2] J.L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs," *IEEE Trans. Signal Processing*, vol. SP-40, pp. 3074-3078, Dec. 1992.

[3] E. Moreau and O. Macchi, "New self-adaptive algorithms for source separation based on contrast functions," in *Proc. IEEE Signal Proc. Workshop on Higher-Order Statistics*, South Lake Tahoe, CA, pp. 215-219, June 1993.

[4] L.Tong, Y. Inouye and R. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. SP-41, pp. 2461-2470, July 1993.

[5] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-F, Radar and Signal Processing*, vol. 140, pp. 362-370, Dec. 1993.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, No. 3, pp. 287-314, 1994.

[7] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Trans. Signal Processing*, vol. SP-42, pp. 2158-2168, Aug. 1994.

[8] A. Mansour and C. Jutten, "Fourth-order criteria for blind source separation," *IEEE Trans. Signal Processing*, vol. SP-43, pp. 2022-2025, Aug. 1995.

[9] A. Swami, G.B. Giannakis and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Proc.*, vol. SP-42, pp. 898-914, April 1994.

[10] L. Tong, Y. Inouye and R. Liu, "A finite-step global convergence algorithm for parameter estimation of multichannel MA processes," *IEEE Trans. Signal Processing*, vol. SP-40, pp. 2547-2558, Oct. 1992.

[11] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input-multi-output systems driven by colored inputs," *IEEE Trans. Signal Proc.*, vol. SP-45, pp. 1543-1552, June 1997.

[12] X.-R. Cao and R.-W. Liu, "General approach to blind source separation," *IEEE Trans. Signal Proc.*, vol. SP-44, pp. 562-571, March 1996.

[13] J.K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 658-672, March 1997.

[14] J.R. Treichler and M.G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, pp. 420-431, April 1985.

[15] M. Rosenblatt, *Stationary Sequences and Random Fields.* Birkhäuser: Boston, 1985.

[16] J.K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Transactions on Signal Processing*, vol. SP-45, pp. 268-271, Jan. 1997.

[17] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. Intern. Conf. Commun.*, pp. 1160-1164, Geneva, Switzerland, June 1993.

[18] J.K. Tugnait, "Parameter identifiability of multichannel ARMA models of linear non-Gaussian signals via cumulant matching," *IEEE Transactions on Signal Processing*, vol. SP-43, pp. 3067-3069, Dec. 1995.

[19] Y. Inouye and B. Sako, "Identifiability of multichannel linear systems driven by colored inputs and its application to blind signal separation," in *Proc. ISCAS-94*, pp. 57-60, vol. 5, 1994.

[20] Y. Inouye and K. Hirano, "Blind identification of linear multi-input-multi-output systems driven by colored inputs with applications to blind signal separation," in *Proc. 34th IEEE Conf. Decision & Control*, pp. 715-720, New Orleans, LA, Dec. 1995.

[21] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals using adaptive FIR lossless filters," in *Proc. EUSIPCO 92*, pp. 1061-1064, Brussels, Belgium, Aug. 1992.

[22] J.K. Tugnait, "Spatio-temporal signal processing for blind separation of multichannel signals," in *Digital Signal Processing Technology*, Joseph Picone, Editor, *Proc. SPIE 2750*, pp. 88-103, 1996. [Proceedings of the SPIE Conf. held in Orlando, FL, April 10-11, 1996.]

[23] J.K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals," in *Proc. Eighth IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 312-315, Corfu, Greece, June 24-26, 1996.

[24] E. Weinstein, A.V. Oppenheim, M. Feder and J.R. Buck, "Iterative and sequential algorithms for multisensor signal enhancement," *IEEE Trans. Signal Proc.*, vol. SP-42, pp. 846-859, April 1994.

[25] U. Lindgren, T. Wigren and H. Broman, "On local convergence of a class of blind separation algorithms," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 3054-3058, Dec. 1995.

[26] D.H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. IEE*, vol. 130, pts. F & H, pp. 11-16, Feb. 1983.

[27] H.V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag: New York, 1988.

[28] S.U.H. Qureshi, "Adjustment of the position of the reference tap of an adaptive equalizer," *IEEE Trans. Commun.*, vol. COM-21, pp. 1046-1052, Sept. 1973.

[29] O. Macchi and E. Moreau, "Self-adaptive source separation, part I: convergence analysis of a direct linear network controlled by Herault-Jutten algorithm," *IEEE Trans. Signal Proc.*, vol. SP-45, pp. 918-926, April 1997.

[30] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures," in *Proc. 1996 ICASSP*, pp. 2940-2943, Atlanta, GA, May 7-10, 1996.

[31] H.L. Nguyen and C. Jutten, "Blind source separation for convolutive mixtures," *Signal Processing*, vol. 45, pp. 209-229, 1995.

[32] C. Serviere, "Blind source separation of convolutive mixtures," in *Proc. Eighth IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 316-319, Corfu, Greece, June 24-26, 1996.

[33] J.F. Cardoso and B.H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Proc.*, vol. SP-44, pp. 3017-3030, Dec. 1996.

[34] S. Haykin, *Adaptive Filter Theory*, 3rd Ed. Prentice-Hall: Upper Saddle River, NJ, 1996.

[35] A. Mathur et al., "Convergence properties of the multistage constant modulus array for correlated sources," *IEEE Trans. Signal Proc.*, vol. SP-45, pp. 280-286, Jan. 1997.

[36] L. Tong and H.H. Zeng, "Channel surfing reinitialization for the constant modulus algorithm," *IEEE Signal Proc. Letters*, vol. SPL-4, pp. 85-87, March 1997.

Table 1: Example 1: Average SINR after blind separation with record length = 18000 samples. Serial: Algorithm of Sec. 4.3; Parallel: Algorithm of Sec. 4.3 coupled with reinitialization of Sec. 5.2.

SNR	SOURCE 1 (Gaussian mixture)				SOURCE 2 (4-QAM)			
	serial		parallel		serial		parallel	
	(4-37)	MMSE	(4-37)	MMSE	(4-37)	MMSE	(4-37)	MMSE
27 dB	8.656	10.785	8.656	10.785	11.620	12.815	16.618	15.689
20 dB	8454	10.428	8.454	10.428	11.203	12.280	15.333	14.647
14 dB	7.828	9.346	7.828	9.346	9.886	10.695	12.576	12.256
7 dB	5.957	6.594	5.957	6.594	6.548	6.932	7.807	7.718

Table 2: Example 2: Average SINR after blind separation with record length = 18000 samples. Serial: Algorithm of Sec. 4.3; Parallel: Algorithm of Sec. 4.3 coupled with reinitialization of Sec. 5.2.

SNR	SOURCE 1 (Gaussian mixture)				SOURCE 2 (4-QAM)			
	serial		parallel		serial		parallel	
	(4-37)	MMSE	(4-37)	MMSE	(4-37)	MMSE	(4-37)	MMSE
25.2 dB	8.653	10.667	8.653	10.667	11.621	12.647	16.123	15.271
18.2 dB	8.447	10.317	8.447	10.317	11.198	12.134	15.078	14.351
12.2 dB	7.807	9.253	7.807	9.253	9.876	10.591	12.445	12.070
5.2 dB	5.893	6.511	5.893	6.511	6.505	6.862	7.746	7.626

Fig. 1. Example 1

(Algorithm of Sec. 4.3, no reinitialization)

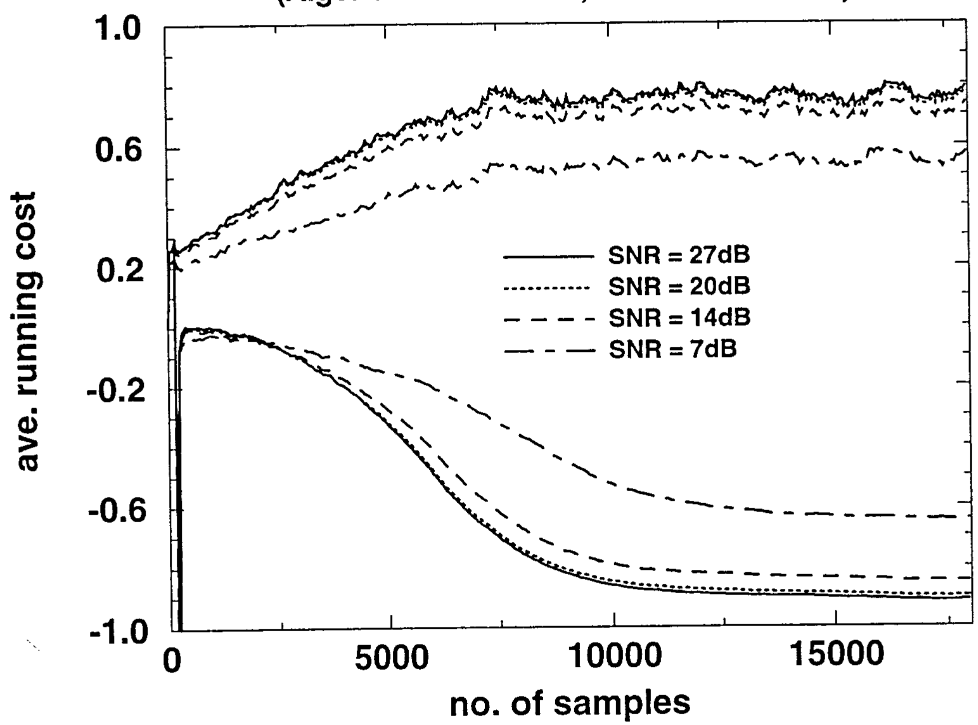


Fig. 2. Example 1

(Algorithm of Sec. 4.3 with reinitialization of Sec. 5.2)

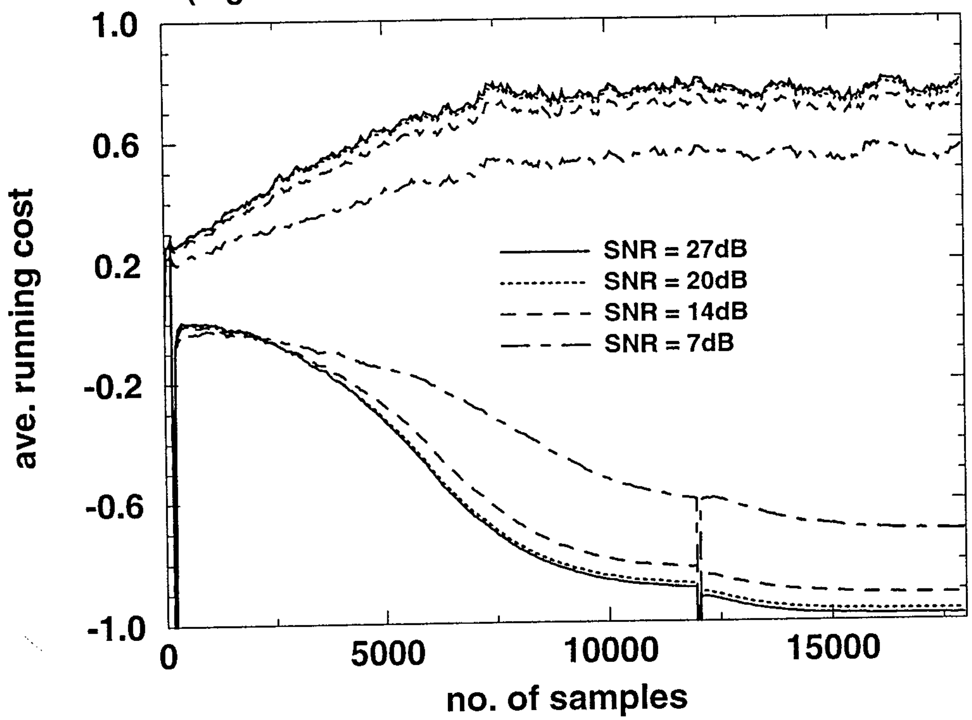


Fig. 3. Example 2

(Algorithm of Sec. 4.3, no reinitialization)

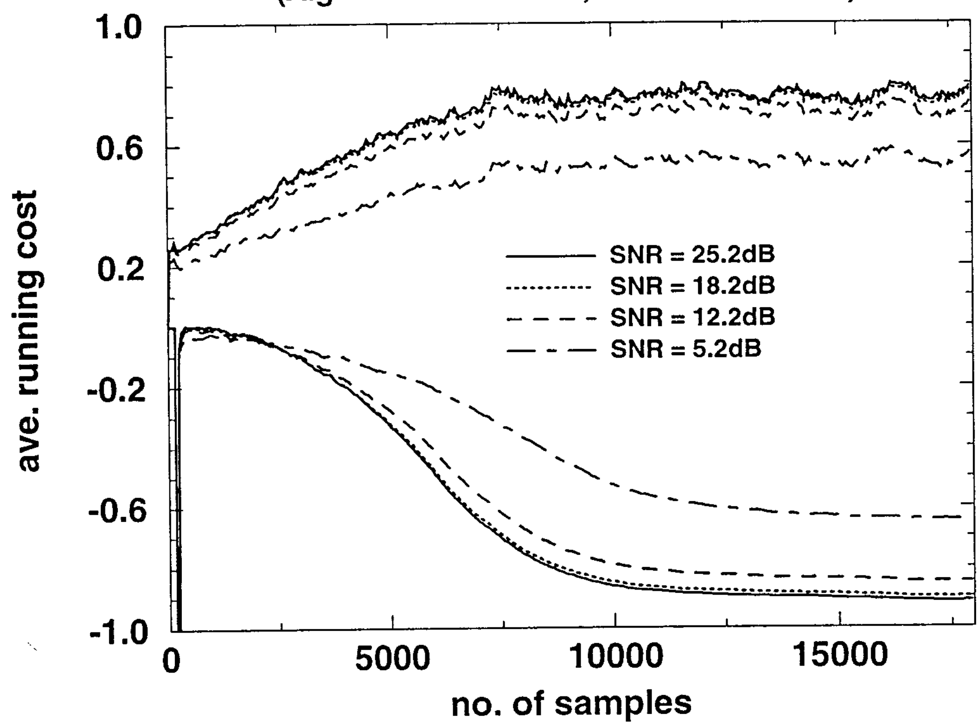
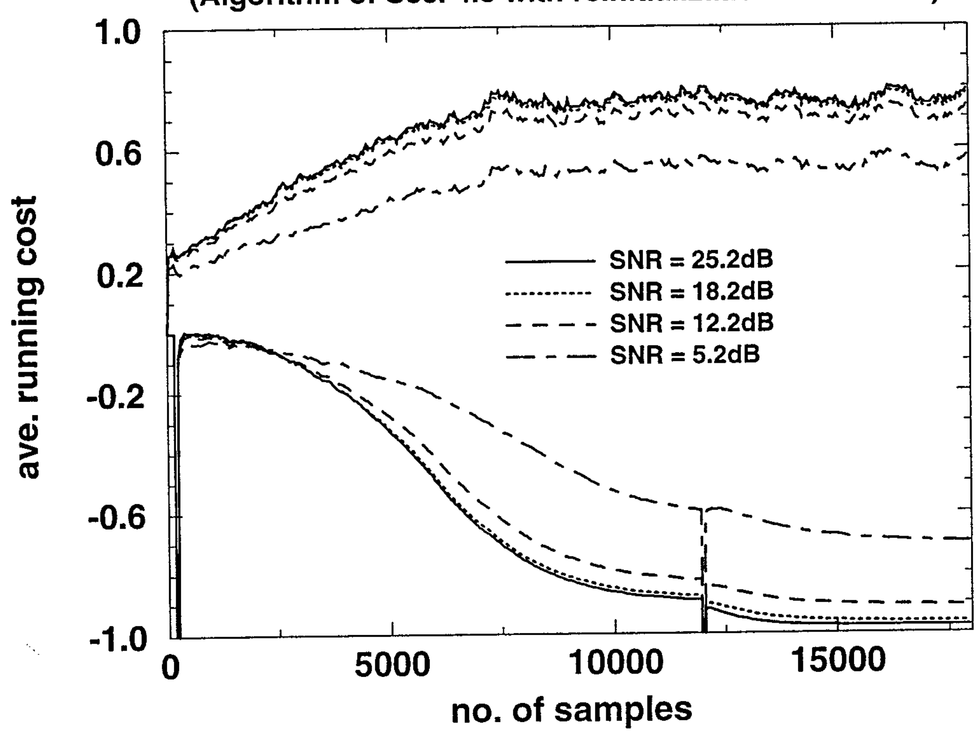


Fig. 4. Example 2

(Algorithm of Sec. 4.3 with reinitialization of Sec. 5.2)



[6]

Blind Adaptive Spatio-Temporal Multiuser Signal Separation And Interference Suppression For Frequency Selective Multipath Channels[1]

Jitendra K. Tugnait

Department of Electrical Engineering

Auburn University, Auburn, AL 36849, USA

Tel.: (334)844-1846 FAX: (334)844-1809

Email: tugnait@eng.auburn.edu

Abstract

This paper is concerned with the problem of blind adaptive deconvolution of multiple communications signals and estimation of the matrix impulse response function of the underlying multiple-input multiple-output system given only the measurements of the vector output of the system. The multiple signals are received at an antenna array in the presence of both interuser as well as intersymbol interference. Recently a source-iterative, inverse filter criteria based approach was developed using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach was input-iterative, i.e., the inputs were extracted and removed one-by-one. The matrix impulse response was then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper an adaptive implementation of the above approach is developed using a stochastic gradient approach. Computer simulation examples are presented to illustrate the proposed approach.

Keywords: Spatio-temporal processing, signal separation, space division multiple access, multi-input multi-output channels, co-channel interference suppression, space diversity, time diversity (fractional sampling)

[1]This work was supported by the National Science Foundation under Grant MIP-9312559 and by the Office of Naval Research under Grant N00014-97-1-0822.

1 Introduction

Multiuser wireless communications systems have attracted considerable attention in recent years. Because of limited frequency spectrum allocated, approaches that lead to increased spectrum efficiency are of much interest. One promising concept is to use antenna arrays to discriminate among signals that have distinct spatial signatures (multipaths etc.) – SDMA (space division multiple access) [9]-[11]. This allows several sources using the same carrier frequency to also use the same time slot in a given cell thereby increasing the system capacity. In this paper we consider the problem of separating multiple signals (including possibly non-digital communications interferences) received at an antenna array. The signals are allowed to undergo multipath propagation where the delay spreads are not necessarily negligible.

The baseband-equivalent mathematical model for the problem under consideration is that of a multiple-input multiple-output (MIMO) system. Such modeling of digital communication systems has received considerable attention recently in a variety of contexts (other than SDMA) [13]-[16]. A major limiting factor in high data rate ($\geq$ 800 kbps) digital subscriber lines (DSL) using twisted-pair wires is crosstalk between twisted pairs in close physical proximity [14]. In [14] the entire cable has been treated as a single MIMO channel with the crosstalk characterized by the matrix impulse response of the channel, rather than as additive noise. [14] is concerned with design of linear equalizers for suppression of near- and far-end crosstalk assuming complete knowledge of the MIMO channel matrix transfer function. In multi-track digital magnetic recording [17] MIMO representation is needed to represent crosstalk arising from adjacent tracks. Even in a single-track situation, MIMO models may arise because of vector stationary process modeling of scalar cyclostationary signals [13]. For instance, many run-length limited codes used in magnetic recording give rise to cyclostationary sequences [13]. Other applications include dually polarized radio channels [18] and multisensor sonar/radar systems [19].

Some of the recent work on MIMO channels has been concerned with design of transmitter and receiver filters [14] and [15], and MIMO equalizers for suppression of intersymbol interference (ISI), cochannel and adjacent channel interferences (CCI and ACI) [13] and [16]. In these contributions complete knowledge of the MIMO transfer function is assumed to be available. As noted in [16] (see also [17]), the equalizers can be adapted using LMS (least-mean squares) or other algorithms based on minimizing the mean square error between the actual response (of the equalizer) and the

desired response which is typically supplied by a training sequence. In case of MIMO channels, the training sequence has to be a vector sequence which, in turn, implies that cochannel and adjacent channel interferences must also cooperate in furnishing training sequences. Clearly this is unrealistic. In other situations even the transmitter of the desired signal may not be able to transmit a training sequence. This leads to the desirability of adaptive design of MIMO equalizers and channel estimators *in the absence of any training sequences*: **blind channel estimation and equalization.** This paper is concerned with exactly this problem.

Past work on blind equalization and/or channel estimation has been overwhelmingly concentrated upon SISO systems (single signal single channel scenario with baud-rate sampled data). First blind adaptive equalizer was proposed by Sato [20]. This work was followed by generalizations due to Godard [21] and to Benveniste et al. [22]. The CMA (constant modulus algorithm) [23] is a special case of and an alternative interpretation of the Godard family of equalizers. Other contributions to SISO systems blind equalization problem include [25]-[26] (and references therein). The communications channels are, in general, nonminimum-phase; hence, the second-order statistics of the baud-rate-sampled stationary signals are inadequate for blind channel identification [22],[24],[27]; that is, in general, conventional LMS (least-mean square) scheme will not work in a blind setting [27]. In [1]-[2] (and others [33]) it is proposed to use fractional sampling and to exploit the second-order statistics of the fractionally sampled data which are cyclostationary. It has been shown in [28] that for a class of multipath channels, the approaches of [1]-[2] will be unable to correctly identify the underlying channel transfer function. In particular, this class includes all multipath channels consisting of time delays that are integer multiples of the symbol duration. No such problems arise if higher-order statistics of the data are also exploited [28].

Prior work on blind equalization and/or channel estimation for truly MIMO systems (more than one information sequence) has been far less extensive. References [3]-[7], [9]-[12], [23], [29] and [30]-[33] (and references therein) have considered this problem in the communications context. In an interesting paper [12] it has been pointed out that given a complex(-valued) MIMO channel and consequently a complex equalizer, but with real-valued signals (sources such as M-ary PAM), the CMA/Godard cost functions (and their variations) employed in [7], [10], [23], [29], [30] and [32] will have some undesirable global minima in that the real and imaginary parts of each equalizer output after convergence, may correspond to different user signals. This would then necessitate

further processing to check for and correct such misconvergence. In a mixed source scenario where not all users have the same alphabet (e.g. 2-PAM and 4-QAM), it is not clear how one would detect such a misconvergence. We note that such a misconvergence can not occur for the cost function considered in this paper (see [5] for a discussion of the convergence points of the cost used). The approaches of [7], [10], [23], [30] and [32] are restricted to symmetric sources with negative fourth cumulants. In this paper we allow the sources to be asymmetric and they can have negative or positive fourth cumulants.

Finite alphabet property (all sources have the same alphabet) has been used in [9] where delay spread has been assumed to be negligible. We assume neither in this paper: different users are allowed to have different alphabets and moreover, some "users" may actually represent non-Gaussian non-communications interferences. Moreover, we consider multipath propagation with non-negligible delay spreads. In [3] a subspace approach approach has been used whereas in [31] a subspace approach coupled with the finite alphabet property (with known and identical alphabets) has been proposed. The subspace approaches of [3] and [31] require the MIMO transfer function $\mathcal{F}(z)$ (see (2-2)) to have full rank for every z including $z = \infty$ but excluding $z = 0$ whereas in this paper we only require $\mathcal{F}(z)$ to have full rank for $|z| = 1$; see Theorem 1 in Sec. 3.

In [4],[5] an iterative, inverse filter criteria based approach has been developed for deconvolution of multichannel non-Gaussian processes using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach is input-iterative, i.e., the inputs are extracted and removed one-by-one. The matrix impulse response is then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper we develop a stochastic gradient-based "recursification" of all of the batch optimization steps in [4],[5]. An interesting input-iterative adaptive approach using prewhitened observations and the fourth-order cumulant of the inverse-filtered data at zero-lag has been considered in [34] and [35]. The inverse filter is constrained to have a lossless filter structure which is realized using a lossless lattice filter. Such a restriction can lead to ill-conditioning of the algorithm of [34] as one iteratively extracts input sequences. A fix to this is proposed in [35] but it works only for the two-input case. Refs. [34] and [35] are restricted to 'square' systems: number of inputs (M) equal to the number of outputs (N), whereas in this paper we allow $N \geq M$, a common occurrence in array processing. Moreover, in this paper we perform no prewhitening, rather we operate directly on the given measurements. A consequence of this is that the ill-conditioning of

[34],[35] referred to above does not occur in our approach. Refs. [34] and [35] are restricted to real-valued data whereas we also consider complex-valued observations.

The paper is organized as follows. In Sec. 2 the precise model assumptions are presented. The inverse-filter criteria-based approach of [5], the underlying identifiability results and the iterative source separation solution of [5] are briefly discussed in Sec. 3. In Sec. 4 we develop a stochastic gradient-based "recursification" of all of the batch optimization steps discussed in Sec. 3. Computer simulation examples are presented in Sec. 5.

2 Problem Statement and Assumptions

Consider a discrete-time MIMO system, possibly complex-valued, with N outputs and M inputs. The i-th component of the output at time k is given by

$$y_i(k) = \sum_{j=1}^{M} \mathcal{F}_{ij}(z) w_j(k) + n_i(k), \quad i = 1, 2, \cdots, N, \tag{2-1}$$

$$\Longrightarrow \quad \mathrm{y}(k) = \mathcal{F}(z)\mathrm{w}(k) + \mathrm{n}(k), \tag{2-2}$$

where $\mathrm{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathrm{w}(k)$ and $\mathrm{n}(k)$, z^{-1} denote both the backward-shift operator (i.e., $z^{-1}w(k) = w(k-1)$, etc.) as well as the complex variable z in the $\mathcal{Z}-$transform, $w_j(k)$ is the j-th input at sampling time k, $y_i(k)$ is the i-th output, $n_i(k)$ is the additive Gaussian measurement noise independent of $\{\mathrm{w}(k)\}$, and

$$\mathcal{F}_{ij}(z) := \sum_{l=-\infty}^{\infty} f_{ij}(l) z^{-l} \tag{2-3}$$

is the scalar transfer function with $w_j(k)$ as the input and $y_i(k)$ as the output. The MIMO transfer function is $\mathcal{F}(z)$ with $ij-$th element $\mathcal{F}_{ij}(z)$. The model (2-1)-(2-2) is the space-time baseband-equivalent channel model used by several authors (e.g. [3]-[7], [10]-[12] and references therein). The above model could be the result of baud-rate sampling of continuous-time signals at N sensors, or it could be the result of oversampling (fractional sampling) at fewer than N sensors [1]-[3].

The following assumptions are made concerning the system model:

(AS1) The vector sequence $\{\mathrm{w}(k)\}$ is zero-mean, temporally i.i.d. (independent and identically distributed) and spatially independent, i.e., various components of $\mathrm{w}(k)$ are independent of each other but not necessarily identically distributed. Assume that the fourth-order

cumulant (see (3-1) later) of all the components of $\mathbf{w}(k)$ are nonzero but not necessarily negative.

(AS2) If it is an infinite impulse response (IIR) model. then (2-2) is assumed to be the result of a finite-dimensional multichannel ARMA model such that the model matrix impulse response function is exponentially stable, i.e., $\| [f_{ij}(l)] \| < a\beta^{|l|}$ for some $0 < a < \infty$ and $0 < \beta < 1$ where $[f_{ij}(l)]$ denotes a matrix with its ij-th element as $f_{ij}(l)$.

(AS3) $N \geq M$, i.e. at least as many outputs as inputs.

(AS4) $\text{Rank}\{\mathcal{F}(z)\} = M$ for any $|z| = 1$.

Notice that we allow the fourth-order cumulants of some components of $\mathbf{w}(k)$ to be positive. This implies that not all of the signals impinging upon the array are necessarily digital communications signals. Moreover, we do not require $E\{w_j^2(k)\} = 0$ if the component $w_j(k)$ has negative fourth cumulant; this is in contrast to the CMA/Godard algorithm-based approaches where we also must have $E\{w_j^2(k)\} = 0$ in addition to negative fourth cumulant of $w_j(k)$. The objective is to recover $w_j(k)\ \forall j$. For non-communications signals (interferences), one may be interested in analyzing the sources (direction, for instance) of such interference. One may not know in advance the number of such interfering sources, consequently, the existing methods (such as [9] and [31]) that exploit the finite alphabet property of the digital communications signals to simultaneously extract all of the sources will not work for the stated problem.

As noted earlier in Sec. 1, it has been pointed out in [12] that for complex MIMO channel-equalizer cascades, but with real-valued sources, the CMA/Godard costs will have some undesirable global minima. "The real and imaginary parts of each equalizer output after convergence, may correspond to different user signals" [12]. It has been shown in [12] that the reason for this is that such real-valued signals are asymmetric (i.e. $E\{w_j^2(k)\} \neq 0$). Such a misconvergence can not occur for the cost function (3-3) considered in this paper [5].

3 An Iterative Solution Based on Inverse-Filter Criteria

In [4],[5] an iterative, inverse filter criteria based approach has been developed for deconvolution of multichannel non-Gaussian processes using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach is input-iterative, i.e., the inputs are extracted and removed

one-by-one. The matrix impulse response is then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper we develop a stochastic gradient-based "recursification" of all of the batch optimization steps in [4],[5]. In this section we briefly discuss the batch (non-recursive) approach of [4],[5]; its adaptive version is developed in Sec. 4.

Let $\mathrm{CUM}_4(w)$ denote the fourth-order cumulant of a complex-valued scalar zero-mean random variable w, defined as

$$\mathrm{CUM}_4(w) := \mathrm{cum}_4\{w, w^*, w, w^*\} = E\{|w|^4\} - 2[E\{|w|^2\}]^2 - |E\{w^2\}|^2 \tag{3-1}$$

where $*$ denotes complex conjugation. We will use the notation $\gamma_{4wi} = \mathrm{CUM}_4(w_i(k))$ and $\sigma_{wi}^2 = E\{|w_i(k)|^2\}$. Consider an $1 \times N$ row-vector polynomial equalizer (filter) $\mathcal{C}^T(z)$, with its j-th entry denoted by $\mathcal{C}_j(z)$, operating on the data vector $\mathbf{y}(k)$. Let the equalizer output be denoted by $e(k)$:

$$e(k) = \sum_{i=1}^{N} \mathcal{C}_i(z) y_i(k). \tag{3-2}$$

Following [4] consider maximization of the cost

$$J := \frac{|\mathrm{CUM}_4(e(k))|}{[E\{|e(k)|^2\}]^2} \tag{3-3}$$

for designing a linear equalizer to recover one of the inputs. It is shown [4] that when (3-3) is maximized w.r.t. $\mathcal{C}(z)$, then (3-2) reduces to

$$e(k) = d w_{j_0}(k - k_0), \tag{3-4}$$

where d is some complex constant, k_0 is some integer, j_0 indexes some input out of the given M inputs, i.e., the equalizer output is a possibly scaled and shifted version of one of the system inputs. It has been established in [5] that under **(AS1)**-**(AS4)** and no noise, such a solution exists and if doubly-infinite equalizers are used, then all locally stable stationary points of the given cost w.r.t. the equalizer coefficients are also characterized by solutions such as (3-4).

An source-iterative solution is given by:

Step 1. Maximize (3-3) w.r.t. the equalizer $\mathcal{C}(z)$ to obtain (3-4). Let

$$\gamma_{4j_0} = \mathrm{CUM}_4(e(k)) = \mathrm{CUM}_4(d w_{j_0}(k)). \tag{3-5}$$

Step 2. Cross-correlate $\{e(k)\}$ (of (3-4)) with the given data (2-2) and define a possibly scaled and shifted estimate of $f_{ij_0}(\tau)$ as

$$\widehat{f}_{ij_0}(\tau) := \frac{E\{y_i(k)e^*(k-\tau)\}}{E\{|e(k)|^2\}} \tag{3-6}$$

where $F_{ij}(z) = \sum_{l=-\infty}^{\infty} f_{ij}(l) z^{-l}$. Consider now the reconstructed contribution of $e(k)$ to the data $y_i(k)$ $(i = 1, 2, \cdots, N)$, denoted by $\widehat{y}_{i,j_0}(k)$:

$$\widehat{y}_{i,j_0}(k) := \sum_l \widehat{f}_{ij_0}(l) e(k-l). \tag{3-7}$$

Step 3. Remove the above contribution from the data to define the outputs of a MIMO system with N outputs and $M-1$ inputs. These are given by

$$y_i'(k) \ := \ y_i(k) - \widehat{y}_{i,j_0}(k). \tag{3-8}$$

Step 4. If $M > 1$, set $M \leftarrow M-1$, $y_i(k) \leftarrow y_i'(k)$, and go back to **Step 1**, else quit.

In practice, all the expectations in (3-6) are replaced with their sample averages over appropriate data records.

It has been shown in [4],[5] that

$$\widehat{y}_{i,j_0}(k) = \sum_l f_{ij_0}(l) w_{j_0}(k-l), \tag{3-9}$$

i.e., we have decomposed the observations at the various sensors into its independent components: $\widehat{y}_{i,j_0}(k)$ in (3-9) represents the contribution of $\{w_{j_0}(k)\}$ to the i-th sensor achieving **blind signal separation**. This aspect may be useful in isolation and analysis of non-communication interfering signals.

Remark 1. It has been shown in [5] that under the conditions (**AS1**)-(**AS4**) and no noise, the proposed iterative approach is capable of blind identification of a MIMO transfer function $\mathcal{F}(z)$ up to a time-shift, a scaling and a permutation matrix provided that we allow doubly-infinite equalizers. That is, given $\mathcal{F}(z)$, we end up with a $\mathcal{A}(z)$ where the two are related via

$$\mathcal{A}(z) \ = \ \mathcal{F}(z)\mathbf{DAP} \tag{3-10}$$

where $\mathbf{D}$ is an $M \times M$ "time-shift" diagonal matrix with diagonal entries such as z^{-k_0} (recall (3-4)), $\mathbf{A}$ is an $M \times M$ diagonal scaling matrix, and $\mathbf{P}$ is an $M \times M$ permutation matrix. The following result has been proved in [5]

Theorem 1[5]: Given the model (2-2) such that $\mathbf{n}(k) \equiv 0$ and given the true 4th-order and 2nd-order cumulant functions of the model output $\{\mathbf{y}(k)\}$ such that conditions (**AS1**)-(**AS4**) hold true. Suppose that doubly infinite equalizers are used in steps 1–4 of the iterative procedure of Sec. 3. Then this procedure yields a transfer function $\mathcal{A}(z)$ satisfying (3-10). • $\square$

Remark 2. The results of [4],[5] are based upon the use of doubly-infinite inverse filters. If we

assume that $\mathcal{F}(z)$ has finite impulse response (FIR) and rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), then finite length inverse filters suffice. For an analysis and further elaborations, see [6] and [7] where a Godard cost function is considered but the results of [6] and [7] can be easily modified to apply to the cost (3-3) and the basic conclusions remain unchanged. The following result follows from [5] and [7].

Theorem 2: Given the FIR model (2-2) such that $\mathbf{n}(k) \equiv 0$ and conditions **(AS1)** and **(AS4)** hold true. Suppose that steps 1–4 of the iterative procedure of Sec. 3 are used and the record length tends to infinity. Then this procedure yields a transfer function $\mathcal{A}(z)$ satisfying (3-10) if one of the following holds true:

(A) Rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), and doubly-infinite equalizers are used.

(B) Rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), $\mathcal{F}(z)$ is column-reduced and FIR equalizers with length $L_e \geq (2M-1)L_c - 1$ are used where L_c = channel length.

• □

4 Adaptive Algorithm

In this section we develop a stochastic gradient-based "recursification" of all of the batch optimization steps discussed in Sec. 3. Theorems 1 and 2 of Sec. 3 motivate and justify the algorithm developed in this section.

4.1 First Stage Maximization of Normalized Fourth Cumulant

Let the length of the equalizer $\mathcal{C}(z)$ be L_e and let

$$\mathcal{C}_i(z) = \sum_{l=0}^{L_e-1} c_i(l) z^{-l}. \tag{4-1}$$

This allows us to rewrite (3-2) as

$$e(k) = \sum_{i=1}^{N} \sum_{l=0}^{L_e-1} c_i(l) y_i(k-l) = \mathbf{C}^T \mathbf{Y}(k) \tag{4-2}$$

where

$$\mathbf{Y}(k) = \begin{bmatrix} \mathbf{Y}_1^T(k) & \mathbf{Y}_2^T(k) & \cdots & \mathbf{Y}_N^T(k) \end{bmatrix}^T, \tag{4-3}$$

$$Y_i(k) = \left[\begin{array}{cccc} y_i(k) & y_i(k-1) & \cdots & y_i(k-L_e+1) \end{array} \right]^T, \tag{4-4}$$

$$\mathbf{C}(k) = \left[\begin{array}{cccc} C_1^T & C_2^T & \cdots & C_N^T \end{array} \right]^T, \tag{4-5}$$

and

$$C_i = \left[\begin{array}{cccc} c_i(0) & c_i(1) & \cdots & c_i(L_e-1) \end{array} \right]^T. \tag{4-6}$$

Define

$$m_4 \;=\; E\{|e(k)|^4\}, \quad m_2 \;=\; E\{|e(k)|^2\}, \quad \tilde{m}_2 \;=\; E\{e^2(k)\}. \tag{4-7}$$

Then showing explicit dependence upon $\mathbf{C}$, (3-3) may be rewritten as

$$J(\mathbf{C}) \;=\; \mathrm{sgn}(\gamma_4) \left[\frac{m_4 - |\tilde{m}_2|^2}{m_2^2} - 2 \right] \tag{4-8}$$

where

$$\gamma_4 \;=\; m_4 - 2\, m_2^2 - |\tilde{m}_2|^2. \tag{4-9}$$

Let $\nabla_{\mathbf{C}}$ denote a gradient operator (w.r.t. a vector $\mathbf{C}$). We will follow [7] in formally defining the complex derivatives. Then we have

$$\nabla_{\mathbf{C}^*} e(k) = 0 \quad \text{and} \quad \nabla_{\mathbf{C}^*}\, e^*(k) = \mathbf{Y}^*(k). \tag{4-10}$$

Using the above results in (4-7) we have

$$\nabla_{\mathbf{C}^*} m_4 \;=\; 2E\{e^2(k) e^*(k) \mathbf{Y}^*(k)\}, \quad \nabla_{\mathbf{C}^*} m_2 \;=\; E\{e(k)\mathbf{Y}^*(k)\} \tag{4-11}$$

and

$$\nabla_{\mathbf{C}^*} \tilde{m}_2 \;=\; 0, \quad \nabla_{\mathbf{C}^*} \tilde{m}_2 \;=\; 2\, E\{e^*(k)\mathbf{Y}^*(k)\}. \tag{4-12}$$

Using (4-8)-(4-12) and after some simplification, we have

$$\nabla_{\mathbf{C}^*} J(\mathbf{C}) \;=\;$$

$$\frac{2\,\mathrm{sgn}(\gamma_4)}{m_2^3} \left\{ m_2 E\{|e(k)|^2 e(k) \mathbf{Y}^*(k)\} - \tilde{m}_2 m_2 E\{e^*(k)\mathbf{Y}^*(k)\} - [m_4 - |\tilde{m}_2|^2] E\{e(k)\mathbf{Y}^*(k)\} \right\}. \tag{4-13}$$

We will use a stochastic gradient method for recursification of maximization of $J(\mathbf{C})$ using an 'instantaneous' gradient as an estimate of (4-13). Given the estimate $\mathbf{C}(k-1)$ of the tap-gains at time $k-1$, the stochastic gradient method computes the update $\mathbf{C}(k)$ at time k as

$$\tilde{\mathbf{C}}(k) = \mathbf{C}(k-1) + \mu_1 \nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k-1)) \tag{4-14}$$

$$\mathbf{C}(k) = \frac{\tilde{\mathbf{C}}(k)}{||\tilde{\mathbf{C}}(k)||} \tag{4-15}$$

where μ_1 is the update step-size and $\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k-1))$ is an instantaneous gradient of the cost J (w.r.t. $\mathbf{C}^*$) at time k evaluated at $\mathbf{C}(k-1)$. Since the cost J is invariant any scaling of $\mathbf{C}$, we normalize $\mathbf{C}$ in (4-15) to have a unit norm. From (4-13) we have the approximation

$$\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k)) = \operatorname{sgn}(\gamma_{4k}) \frac{2}{m_{2k}^3} \left\{ \left[m_{2k} \left(e^2(k) - \tilde{m}_{2k} \right) e^*(k) - \left(m_{4k} - |\tilde{m}_{2k}|^2 \right) e(k) \right] \mathbf{Y}^*(k) \right\} \tag{4-16}$$

where

$$m_{2k} = (1-\mu_2) m_{2(k-1)} + \mu_2 |e(k)|^2, \tag{4-17}$$

$$\tilde{m}_{2k} = (1-\mu_2) \tilde{m}_{2(k-1)} + \mu_2 e^2(k), \tag{4-18}$$

$$m_{4k} = (1-\mu_2) m_{4(k-1)} + \mu_2 |e(k)|^4, \tag{4-19}$$

$$\gamma_{4k} = m_{4k} - 2\, m_{2k}^2 - |\tilde{m}_{2k}|^2 \tag{4-20}$$

and

$$e(k) = \mathbf{C}^T(k) \mathbf{Y}(k). \tag{4-21}$$

In (4-17)-(4-19) the various quantities represent estimates based upon sample averaging, the (exponential window) memory being controlled by the forgetting factor μ_2 ($0 < \mu_2 < 1$). The initializations for (4-17)-(4-19) are: $m_{20} = m_{40} = \tilde{m}_{20} = 0$.

4.2 First Stage Signal Cancellation

Now we discuss implementation of (3-6) via sample averaging using an exponential window controlled by a forgetting factor μ_3. Define ($L_1, L_2 > 0$)

$$\mathbf{E}(k) = \left[\begin{array}{cccc} e(k+L_1) & e(k+L_1-1) & \cdots & e(k-L_2) \end{array} \right]^T \tag{4-22}$$

and

$$\mathbf{F}_i = \left[\begin{array}{cccc} \tilde{f}_i(-L_1) & \tilde{f}_i(-L_1+1) & \cdots & \tilde{f}_i(L_2) \end{array} \right]^T . \tag{4-23}$$

By Sec. 3, when (4-8) is maximized, $e(k)$ satisfies (3-4) so that for suitable choice of L_1 and L_2, there exists a $j_0 \in \{1, 2, \cdots, M\}$ such that

$$\sum_l f_{ij_0}(l) w_{j_0}(k-l) \;=\; \mathbf{F}_i^T \mathbf{E}(k), \quad i = 1, 2, \cdots, N. \tag{4-24}$$

In order to implement (3-7) and (3-8), we need recursive estimates of $\mathbf{F}_i$. The estimate $\mathbf{F}_i(k)$ of $\mathbf{F}_i$ at time k is provided by

$$\mathbf{F}_i(k) \;=\; \mathbf{R}_i(k)/m_{ee}(k) \tag{4-25}$$

where

$$m_{ee}(k) \;=\; (1-\mu_3) m_{ee}(k-1) \;+\; \mu_3 |e(k)|^2, \tag{4-26}$$

$$\mathbf{R}_i(k) \;=\; (1-\mu_3)\mathbf{R}_i(k-1) \;+\; \mu_3 y_i(k) \mathbf{E}^*(k). \tag{4-27}$$

4.3 Multistage Algorithm

In Secs. 4.1 and 4.2 we discussed the first stage of the algorithm where we have N sensors and M sources. Now we put it all together following the source-iterative solution of Sec. 3 and discuss extraction of M sources including the cancellation of the extracted sources. We will use the superscript (m) to denote the various quantities pertaining to stage m. These have been used previously in Secs. 4.1 and 4.2 without this superscript; for instance, $\mathbf{C}^{(m)}(k)$ now denotes the estimate of the tap-gain vector at time k at stage m, etc.

Initialization:

$$\mathbf{Y}^{(1)}(k) \;=\; \quad \text{as in (4-3)} \tag{4-28}$$

DO FOR $m = 1, 2, \cdots, M$**:**

$$\tilde{\mathbf{C}}^{(m)}(k) \;=\; \mathbf{C}^{(m)}(k-1) \;+\; \mu_1 \; \nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k-1)) \tag{4-29}$$

$$\mathbf{C}^{(m)}(k) \;=\; \frac{\tilde{\mathbf{C}}^{(m)}(k)}{\|\tilde{\mathbf{C}}^{(m)}(k)\|} \tag{4-30}$$

where

$$\nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k)) = \operatorname{sgn}(\gamma_{4k}^{(m)}) \frac{2}{m_{2k}^{(m)3}} \left\{ \left[m_{2k}^{(m)} \left(e^{(m)2}(k) - \tilde{m}_{2k}^{(m)} \right) e^{(m)*}(k) \right. \right.$$

$$\left. \left. - \left(m_{4k}^{(m)} - |\tilde{m}_{2k}^{(m)}|^2 \right) e^{(m)}(k) \right] \mathbf{Y}^{(m)*}(k) \right\}, \tag{4-31}$$

$$m_{2k}^{(m)} = (1-\mu_2) m_{2(k-1)}^{(m)} + \mu_2 |e^{(m)}(k)|^2, \tag{4-32}$$

$$\tilde{m}_{2k}^{(m)} = (1-\mu_2) \tilde{m}_{2(k-1)}^{(m)} + \mu_2 e^{(m)2}(k), \tag{4-33}$$

$$m_{4k}^{(m)} = (1-\mu_2) m_{4(k-1)}^{(m)} + \mu_2 |e^{(m)}(k)|^4, \tag{4-34}$$

$$\gamma_{4k}^{(m)} = m_{4k}^{(m)} - 2\, m_{2k}^{(m)2} - |\tilde{m}_{2k}^{(m)}|^2 \tag{4-35}$$

and

$$e^{(m)}(k) = \mathbf{C}^{(m)T}(k) \mathbf{Y}^{(m)}(k). \tag{4-36}$$

Set

$$\widehat{y}_i^{(m)}(k) = \mathbf{F}_i^{(m)T}(k) \mathbf{E}^{(m)}(k) \tag{4-37}$$

where $\widehat{y}_i^{(m)}(k)$ represents (cf. (3-7)) the contribution of the extracted source at the $m-$th stage to the measurement at time k at the $i-$th sensor, and where

$$\mathbf{F}_i^{(m)}(k) = \mathbf{R}_i^{(m)}(k) / m_{ee}^{(m)}(k), \tag{4-38}$$

$$m_{ee}^{(m)}(k) = (1-\mu_3) m_{ee}^{(m)}(k-1) + \mu_3 |e^{(m)}(k)|^2, \tag{4-39}$$

$$\mathbf{E}^{(m)}(k) = \begin{bmatrix} e^{(m)}(k+L_1) & e^{(m)}(k+L_1-1) & \cdots & e^{(m)}(k-L_2) \end{bmatrix}^T \tag{4-40}$$

and

$$\mathbf{R}_i^{(m)}(k) = (1-\mu_3) \mathbf{R}_i^{(m)}(k-1) + \mu_3 y_i^{(m)}(k) \mathbf{E}^{(m)*}(k). \tag{4-41}$$

Define

$$\widehat{Y}_i^{(m)}(k) = \left[\begin{array}{cccc} \widehat{y}_i^{(m)}(k) & \widehat{y}_i^{(m)}(k-1) & \cdots & \widehat{y}_i^{(m)}(k-L_e+1) \end{array} \right]^T, \tag{4-42}$$

Set

$$\mathrm{Y}^{(m+1)}(k) = \left[\begin{array}{cccc} Y_1^{(m+1)T}(k) & Y_2^{(m+1)T}(k) & \cdots & Y_N^{(m+1)T}(k) \end{array} \right]^T \tag{4-43}$$

where

$$Y_i^{(m+1)}(k) = Y_i^{(m)}(k) - \widehat{Y}_i^{(m)}(k). \tag{4-44}$$

ENDDO

The sequence $\{e^{(m)}(k)\}$ in (4-36) represents the equalized (up to a scale factor and time delay) source at stage m.

Remark 3. If M were unknown the proposed approach will still work in the sense that if M were underestimated, some sources will be missed but the extracted sources will correspond to one of the users (or interferers). If M were overestimated, all the users/interferers will be recovered in addition to some "meaningless junk" outputs in stages $M_0 + 1$ and later where M_0 denotes true number of users. Indeed one can test the 'residuals' (4-44) (see also (3-8)) to check if any significant non-Gaussian components remain in the data before implementing another equalizer in parallel. We do not pursue this aspect in this paper. □

Running Cost. To monitor the convergence of the equalizers in various stages of the algorithm, it is useful to calculate a running cost (4-8) without the sign. Let $J_k^{(m)}$ denote the running cost for the $m-$th stage at time k, given by

$$J_k^{(m)} = \frac{m_{4k}^{(m)} - |\widetilde{m}_{2k}^{(m)}|^2}{m_{2k}^{(m)2}} - 2 \tag{4-45}$$

where

$$m_{2k}^{(m)} = (1-\mu_4)m_{2(k-1)}^{(m)} + \mu_4|e^{(m)}(k)|^2, \tag{4-46}$$

$$\widetilde{m}_{2k}^{(m)} = (1-\mu_4)\widetilde{m}_{2(k-1)}^{(m)} + \mu_4 e^{(m)2}(k), \tag{4-47}$$

and

$$m_{4k}^{(m)} = (1-\mu_4)m_{4(k-1)}^{(m)} + \mu_4|e^{(m)}(k)|^4. \tag{4-48}$$

For all of the simulations presented in Sec. 5, we took $\mu_4 = 0.002$.

5 Simulation Examples

In this section we provide three computer simulation examples to illustrate the proposed blind adaptive algorithm for multiuser signal separation and interference suppression.

5.1 Example 1: 3 BPSK Sources and 7 Sensors

We consider a wireless communications scenario with three ($M = 3$) BPSK user signals arriving at a uniform linear array of $N = 7$ sensors via a frequency selective multipath channel. The array elements are spaced half a wavelength apart. The array measurements are assumed to be sampled at baud rate (for simulation convenience only) with sampling interval T seconds and the three sources have the same baud rate. The relative time delay τ (relative to the first arrival), the angle of arrival θ (in degrees w.r.t. the array broadside) and the relative attenuation factor (amplitude) α for various sources were selected as:

$$w_1 : \quad (\tau, \theta, \alpha) = (0T, 10°, 0.5),\ (1T, 50°, 0.75)$$

$$w_2 : \quad (\tau, \theta, \alpha) = (0T, -20°, 0.5),\ (1T, 45°, 0.45),\ (2T, 15°, -0.65)$$

$$w_3 : \quad (\tau, \theta, \alpha) = (0T, -35°, 0.7),\ (1T, -5°, 0.4).$$

Thus the signals w_1 and w_3 propagate through two paths whereas w_2 passes through three paths. The signals arriving at the array were normalized such that the signal powers for users 1 and 2 are equal, and 3dB higher than the signal power for user 3. Additive white (both temporally and spatially) Gaussian noise was added to the array measurements to achieve a signal-to-noise-ratio (SNR) of 11.55dB (ratio = 100/7) for the strongest user(s). The SNR for a given user $w_j(k)$ is defined as

$$\text{SNR} = \frac{\frac{1}{N}\sum_{i=1}^{N} E\{|\mathcal{F}_{ij}(z)w_j(k)|^2\}}{E\{|n_i(k)|^2\}}.$$

The proposed approach was applied with $M = 3$ equalizers and $M - 1 = 2$ signal cancellers running in parallel, each successive equalizer put in operation after waiting for 200 samples (symbols) w.r.t. the previous stage. The equalizer length was chosen to be 5 taps per sensor ($L_e = 5$ in (4-2)). The initial guess for the tap gains was taken to be center-tap initialization: set $c_i(2) = 1$ for $i = m$ for the $m-$th stage equalizer ($m = 1, 2, 3$) with the remaining tap gains set to zero. The algorithm

step sizes and forgetting factors for each stage m were chosen as: $\mu_1 = 0.003$ in (4-29), $\mu_2 = 0.015$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41). For the running cost (4-45) computation we selected $\mu_4 = 0.002$ in (4-46)-(4-48). The parameters L_1 and L_2 in (4-40) (see also (4-22) and (4-23)) were selected as $L_1 = 15$ and $L_2 = 6$.

Fig. 1 shows the evolution of the average running cost $J_k^{(m)}$ (see (4-45)), averaged over 100 Monte Carlo runs after 'assigning' each equalizer cost to its corresponding extracted source. For BPSK sources the 4th-order normalized cumulant equals -2; therefore, at convergence, the running cost (4-45) should be close to -2. In Fig. 1 we see these values to be around -1.89 which is largely a consequence of noise in the data which affects only the denominator of (4-45) making it larger than it should be. Table 1 shows the signal-to-interference-and-noise ratio (SINR) and the probability of error P_e at the output of each equalizer at selected time instants, averaged over 100 Monte Carlo runs and 3000 symbols. [The equalizer tap gains at the chosen time instants were 'frozen' and used to equalize data of length 3000 symbols in order to calculate SINR and P_e. The equalized data were rotated, scaled and shifted before calculating the two performance measures.] It is seen from Fig. 1 and Table 1 that the proposed approach works well. As noted earlier, [12] has shown that CMA/Godard cost functions will have problems with the user signals considered in this example.

5.2 Example 2: 3 4-QAM Sources and 7 Sensors

This example is the same as Example 1 except that the three user signals are 4-QAM. The other parameters for signal generation and equalization are just as for Example 1 (e.g. user signals 1 and 2 are 3 dB stronger than the user signal 3, etc.). The counterparts to Fig. 1 and Table 1 are now shown in Fig. 2 and Table 2, respectively. For 4-QAM sources the 4th-order normalized cumulant equals -1; therefore, at convergence, the running cost (4-43) should be close to -1. The convergence is now slower, yet the approach still works well. The weaker user signal now takes longer to be extracted.

5.3 Example 3: 2 Mixed Sources and 5 Sensors

In this example we consider a 4-QAM user signal w_1 (4th normalized cumulant as -1) and a non-communications signal w_2 consisting of an i.i.d. complex Gaussian-mixture (independent and identically distributed real and imaginary parts with the real part being $\mathcal{N}(0,1)$ with probability

0.9 and $\mathcal{N}(0,4)$ with probability 0.1) with 4th normalized cumulant as 0.7433 . The multipath channels for the two signals were selected as:

$$w_1 : \quad (\tau, \theta, \alpha) = (0T, 10°, 0.5), (1T, 50°, 0.75)$$

$$w_2 : \quad (\tau, \theta, \alpha) = (0T, -20°, 0.5), (1T, 45°, 0.45), (2T, 15°, -0.65).$$

The two signals have equal power. Additive white Gaussian noise was added to the array measurements to achieve an SNR of 13dB (ratio = 20) for each user signal.

The proposed approach was applied with $M = 2$ equalizers and $M - 1 = 1$ signal cancellers running in parallel, the second equalizer put in operation after waiting for 200 samples. The equalizer length was chosen to be 5 taps per sensor ($L_e = 5$ in (4-2)). The initial guess for the tap gains was taken to be center-tap initialization: set $c_i(2) = 1$ for $i = m$ for the $m-$th stage equalizer ($m = 1, 2$) with the remaining tap gains set to zero. The algorithm step sizes and forgetting factors were chosen as: $\mu_1 = 0.0005$ in (4-29), $\mu_2 = 0.015$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41) when $\gamma_{4k}^{(m)} \leq 0$ (see (4-35)), and $\mu_1 = 0.0001$ in (4-29), $\mu_2 = 0.003$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41) when $\gamma_{4k}^{(m)} > 0$. The parameters L_1 and L_2 in (4-40) were selected as $L_1 = 15$ and $L_2 = 6$. For the running cost (4-45) computation we selected $\mu_4 = 0.002$ in (4-46)-(4-48).

The counterparts to Fig. 1 and Table 1 are now shown in Fig. 3 and Table 3, respectively, where in Table 3 the P_e for signal 2 is omitted (for obvious reasons). The convergence for the source with positive 4th cumulant is quite slow.

6 Conclusions

The problem of separating multiple signals (including possibly non-digital communications interferences) received at an antenna array in a wireless communications system was considered in the absence of any training sequences. The signals are allowed to undergo multipath propagation where the delay spreads are not necessarily negligible. In [4],[5] an iterative, inverse filter criteria based approach has been developed for deconvolution of multichannel non-Gaussian processes using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach is input-iterative, i.e., the inputs are extracted and removed one-by-one. The matrix impulse response is

then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper we developed a stochastic gradient-based recursification of all of the batch optimization steps in [4],[5]. The proposed blind adaptive algorithm was illustrated via three simulation examples involving frequency selective multipath channels.

It has been pointed out in [12] that for complex MIMO channel-equalizer cascades, but with real-valued sources, the CMA/Godard costs will have some undesirable global minima in that the real and imaginary parts of each equalizer output after convergence, may correspond to different user signals. It has been shown in [12] that the reason for this is that such real-valued signals are asymmetric (i.e. $E\{w_j^2(k)\} \neq 0$). Such a misconvergence can not occur for the cost function considered in this paper.

7 References

[1] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, pp. 856-860, Nov. 1991. [Also, *IEEE Trans. Information Theory*, vol. IT-40, pp. 340-349, March 1994.]

[2] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 516-525, Feb. 1995.

[3] D.T.M. Slock, "Blind joint equalization of multiple synchronous mobile users using oversampling," in *Proc. 28th Asilomar Conf. Signals, Systems, Computers*, pp. 1154-1158, Pacific Grove, CA, Oct.-Nov. 1994.

[4] J.K. Tugnait, "Identification of multichannel linear non-Gaussian processes using higher-order statistics," presented at *29th Annual Asilomar Conf. Signals Systems Computers*, Pacific Grove, CA, Oct. 29 - Nov. 1, 1995.

[5] J.K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 658-672, March 1997.

[6] J.K. Tugnait, "Spatio-temporal signal processing for blind separation of multichannel signals," in *Digital Signal Processing Technology*, Joseph Picone, Editor, *Proc. SPIE 2750*, pp. 88-103, 1996. [Proceedings of the SPIE Conf. held in Orlando, FL, April 10-11, 1996.]

[7] J.K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Transactions on Signal Processing*, vol. SP-45, pp. 268-271, Jan.

1997.

[8] D.H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. IEE*, vol. 130, pts. F & H, pp. 11-16, Feb. 1983.

[9] S. Talwar, M. Viberg and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array – Part I: Algorithms," *IEEE Transactions on Signal Processing*, vol. SP-44, pp. 1184-1197, May 1996.

[10] C.B. Papadias and A.J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Letters*, vol. SPL-4, pp. 178-181, June 1997.

[11] J. Xavier and V. Barroso, "Blind SDMA with a signal subspace geometry matched algorithm," in *Proc. IEEE 1997 Intern. Conf. Acoustics, Speech, Signal Processing*, pp. V-3437-3440, April 24-27, 1997, Munich, Germany.

[12] C.B. Papadias, "On the existence of undesirable global minima of Godard equalizers," in *Proc. IEEE 1997 Intern. Conf. Acoustics, Speech, Signal Processing*, pp. V-3941-3944, April 24-27, 1997, Munich, Germany.

[13] A. Duel-Hallen, "Equalizers for multiple input/ multiple output channels and PAM systems with cyclostationary input sequences," *IEEE J. Selected Areas in Commun.*, vol. JSAC-10, pp. 630-639, April 1992.

[14] M.L. Honig, P. Crespo and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE J. Selected Areas in Commun.*, vol. JSAC-10, pp. 614-629, April 1992.

[15] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. COM-42, pp. 3221-3231, Dec. 1994.

[16] B.R. Petersen and D.D. Falconer, "Suppression of adjacent-channel, cochannel, and intersymbol interference by equalizers and linear combiners," *IEEE Trans. Commun.*, vol. COM-42, pp. 3109-3118, Dec. 1994.

[17] J. Cioffi et al., "Adaptive equalization in magnetic-disk storage channels," *IEEE Commun. Magazine*, vol. 28, pp. 14-29, Feb. 1990.

[18] M. Kavehrad and J. Salz, "Cross-polarization cancellation and equalization in digital transmission over dually polarized multipath fading channel," *Bell System Tech. Journal*, vol. 64, pp. 2211-2245, Dec. 1985.

[19] Z. Zvonar, D. Brady and J. Catipovic, "An adaptive linear multiuser receiver for deep water acoustic local area networks," in *Proc. IEEE 1994 Intern. Conf. Acoustics, Speech, Signal Processing*, pp. II-389-392, Adelaide, Australia, April 19-22, 1994.

[20] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Communications*, vol. COM-23, pp. 679-682, June 1975.

[21] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Communications*, vol. COM-28, pp. 1867-1875, Nov. 1980.

[22] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Communications*, vol. COM-32, pp. 871-883, Aug. 1984.

[23] J.R. Treichler and M.G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, pp. 420-431, April 1985.

[24] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Info. Theory*, vol. IT-36, pp. 312-321, March 1990.

[25] O. Shalvi and E. Weinstein, "Universal methods for blind deconvolution," pp. 121-180, in *Blind Deconvolution*, S. Haykin (ed.), Prentice-Hall, 1994

[26] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 2120-2129, Sept. 1995.

[27] J.G. Proakis, *Digital Communications.* 2nd ed., New York, NY: McGraw-Hill, 1989.

[28] J.K. Tugnait, "On blind identifiability of multipath channels using fractional sampling and second-order cyclostationary statistics," in *Proc. GLOBECOM '93*, pp. 2001-2005, Houston, TX, Nov.-Dec. 1993. [Also *IEEE Trans. Information Theory*, vol. IT-41, pp. 308-311, Jan. 1995.]

[29] S. Mayrargue, "A blind spatio-temporal equalizer for a radio-mobile channel using the constant modulus algorithm (CMA)," in *Proc. IEEE 1994 Intern. Conf. Acoustics, Speech, Signal Processing*, pp. II-317-320, Adelaide, Australia, April 19-22, 1994.

[30] A.J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. SP-44, pp. 1131-1195, May 1996.

[31] A.J. van der Veen, S. Talwar and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Transactions on Signal Processing*, vol. SP-45, pp. 173-190, Jan. 1997.

[32] B.J. Sublett, R.P. Gooch and S.H. Goldberg, "Separation and bearing estimation of co-channel signals," in *Proc. IEEE Military Commun. Conf.*, pp. 629-634, Boston, MA, Oct. 1989.

[33] Special Issue, *IEEE Transactions on Signal Processing*, vol. SP-45, Jan. 1997.

[34] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals using adaptive FIR lossless filters," in *Proc. EUSIPCO 92*, pp. 1061-1064, Brussels, Belgium, Aug. 1992.

[35] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. ICC 93*, pp. 1160-1164, Geneva, Switzerland, June 1993.

Table 1: Example 1: Performance measures at selected times: averages over 100 Monte Carlo runs.

# of samples	User 1		User 2		User 3	
	SINR(dB)	P_e	SINR(dB)	P_e	SINR(dB)	P_e
4000	14.89	$< 3 \times 10^{-4}$	14.25	0.0099	14.84	$< 3 \times 10^{-4}$
6000	15.03	$< 3 \times 10^{-4}$	14.81	$< 3 \times 10^{-4}$	15.13	$< 3 \times 10^{-4}$
8000	15.13	$< 3 \times 10^{-4}$	14.94	$< 3 \times 10^{-4}$	15.19	$< 3 \times 10^{-4}$
12000	15.14	$< 3 \times 10^{-4}$	14.97	$< 3 \times 10^{-4}$	15.23	$< 3 \times 10^{-4}$

Table 2: Example 2: Performance measures at selected times: averages over 100 Monte Carlo runs.

# of samples	User 1		User 2		User 3	
	SINR(dB)	P_e	SINR(dB)	P_e	SINR(dB)	P_e
4000	12.36	0.0942	12.78	0.0483	13.75	0.0011
6000	14.36	0.0292	14.54	0.0073	14.70	$< 3 \times 10^{-4}$
8000	15.06	$< 3 \times 10^{-4}$	14.80	$< 3 \times 10^{-4}$	14.93	$< 3 \times 10^{-4}$
12000	15.17	$< 3 \times 10^{-4}$	14.98	$< 3 \times 10^{-4}$	14.98	$< 3 \times 10^{-4}$

Table 3: Example 3: Performance measures at selected times: averages over 100 Monte Carlo runs.

# of samples	User 1		User 2
	SINR(dB)	P_e	SINR(dB)
8000	14.12	0.0072	5.52
12000	14.79	$< 3 \times 10^{-4}$	9.28
16000	15.06	$< 3 \times 10^{-4}$	11.49

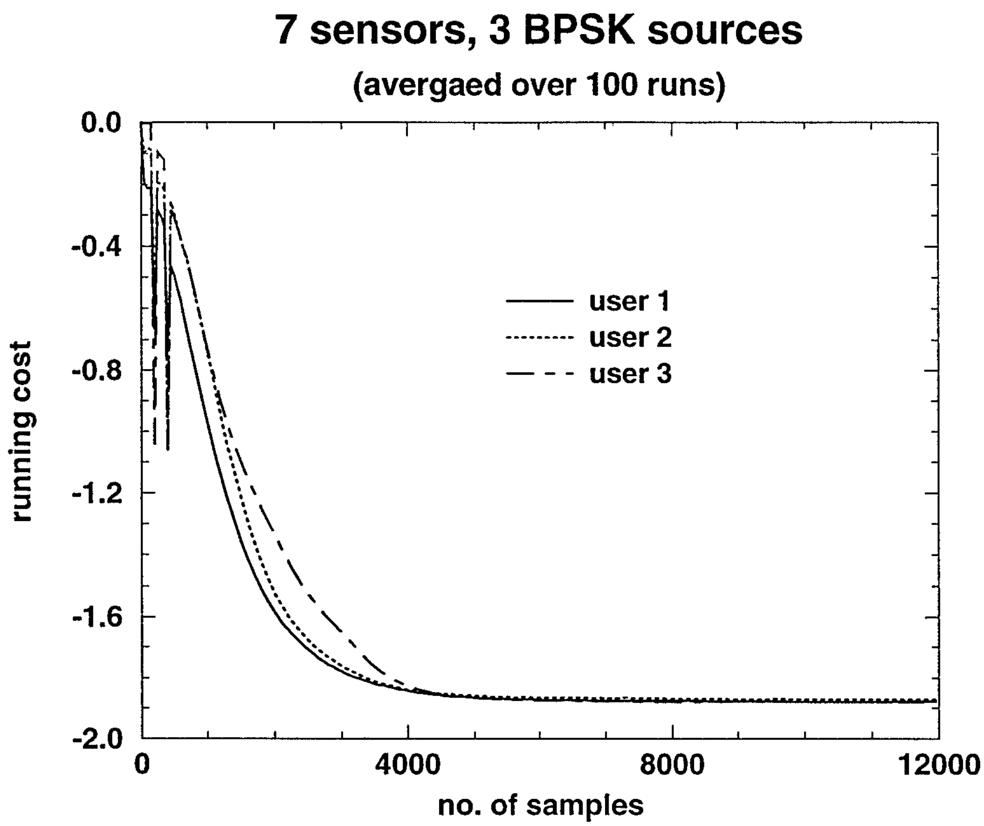


Figure 1: Average running cost for Example 1, averaged over 100 Monte Carlo runs.

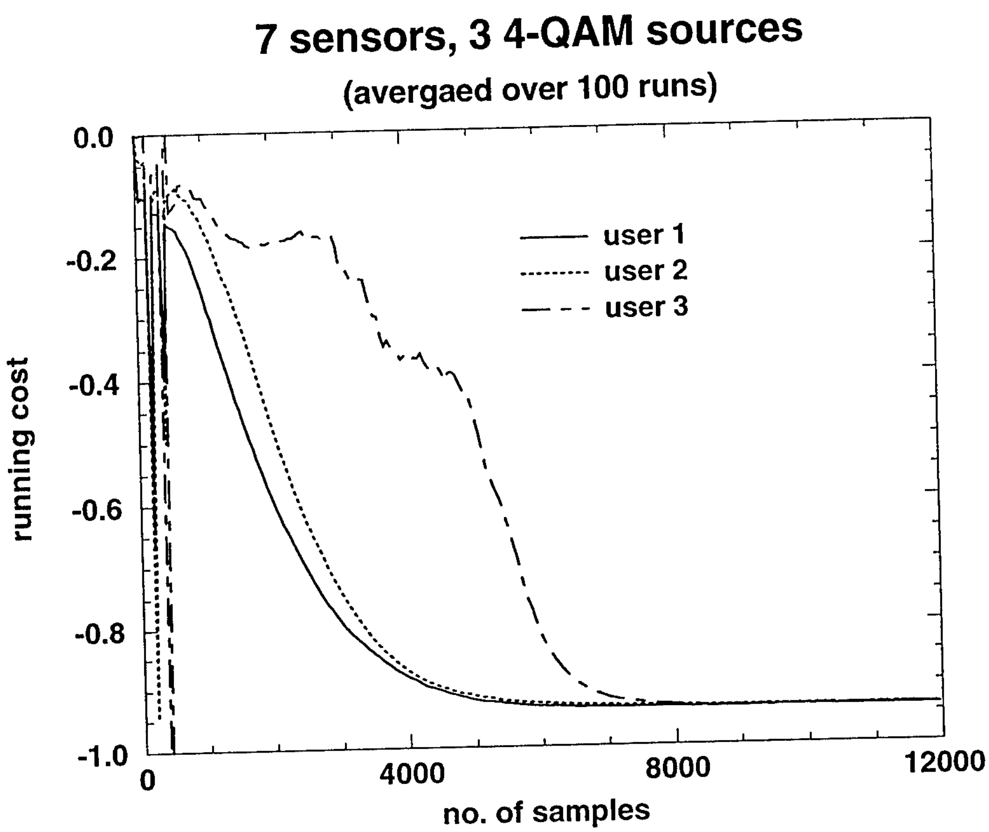


Figure 2: Average running cost for Example 2, averaged over 100 Monte Carlo runs.

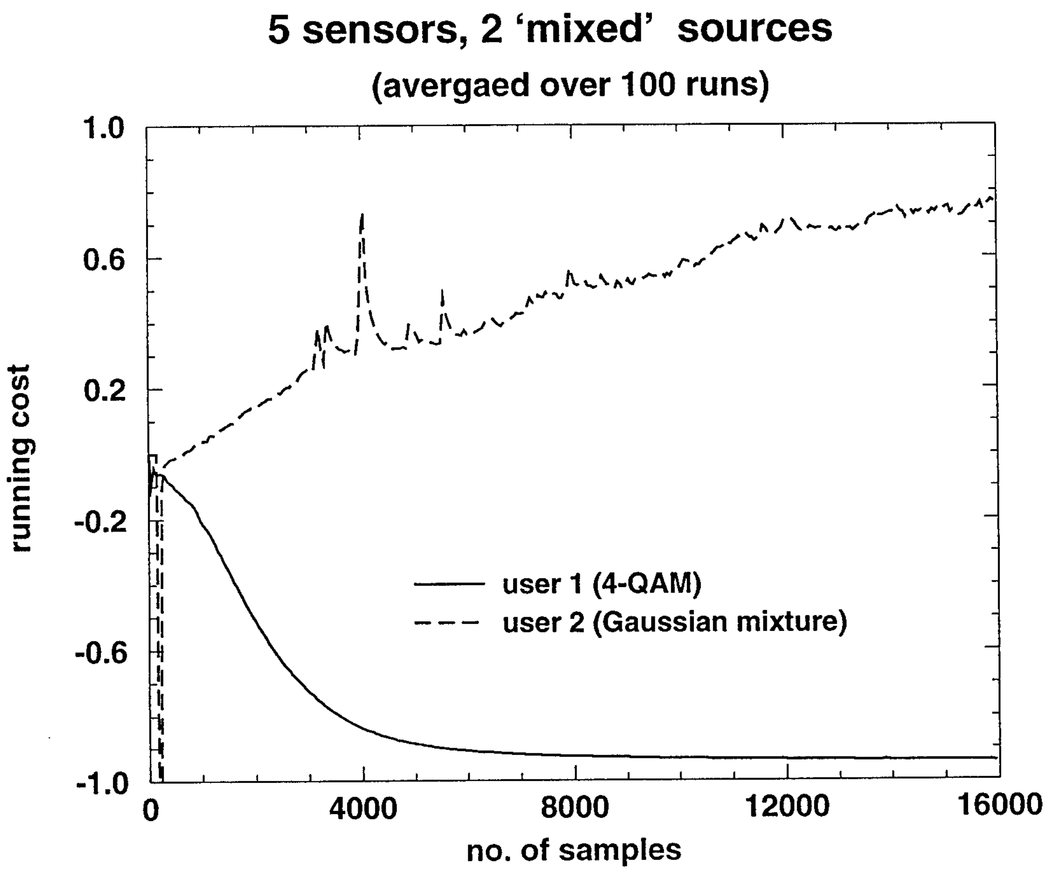


Figure 3: Average running cost for Example 3, averaged over 100 Monte Carlo runs.

[7]

ADAPTIVE BLIND DECONVOLUTION OF M.I.M.O. CHANNELS USING HIGHER-ORDER STATISTICS

Jitendra K. Tugnait

Department of Electrical Engineering
Auburn University, Auburn, Alabama 36849, USA
E-mail: tugnait@eng.auburn.edu

ABSTRACT

This paper is concerned with the problem of adaptive deconvolution and estimation of the matrix impulse response function of a multiple-input multiple-output system given only the measurements of the vector output of the system. The system is assumed to be driven by a spatially and temporally i.i.d. non-Gaussian vector sequence (which is not observed). Recently a iterative, inverse filter criteria based approach was developed using the third-order and/or fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach was input-iterative, i.e., the inputs were extracted and removed one-by-one. The matrix impulse response was then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper an adaptive implementation of the above approach is developed using a stochastic gradient approach. Simulation examples are presented to illustrate the proposed approach.

1. INTRODUCTION

Consider a discrete-time MIMO system, possibly complex-valued, with N outputs and M inputs. The i-th component of the output at time k is given by

$$y_i(k) = \sum_{j=1}^{M} \mathcal{F}_{ij}(z) w_j(k) + n_i(k), \quad i = 1, 2, \cdots, N, \qquad (1-1)$$

$$\Longrightarrow \quad \mathbf{y}(k) = \mathcal{F}(z)\mathbf{w}(k) + \mathbf{n}(k), \qquad (1-2)$$

where $\mathbf{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathbf{w}(k)$ and $\mathbf{n}(k)$, z^{-1} denote both the backward-shift operator (i.e., $z^{-1}w(k) = w(k-1)$, etc.) as well as the complex variable z in the $\mathcal{Z}-$transform, $w_j(k)$ is the j-th input at sampling time k, $y_i(k)$ is the i-th output, $n_i(k)$ is the additive Gaussian measurement noise independent of $\{\mathbf{w}(k)\}$, and $\mathcal{F}_{ij}(z) = \sum_{l=-\infty}^{\infty} f_{ij}(l) z^{-l}$ is the scalar transfer function with $w_j(k)$ as the input and $y_i(k)$ as the output. The MIMO transfer function is $\mathcal{F}(z)$ with $ij-$th element $\mathcal{F}_{ij}(z)$. The model (1-1)-(1-2) is the space-time baseband-equivalent channel model used by several authors (e.g. [3]-[7], [9]-[10] and references therein). The above model could be the result of baud-rate sampling of continuous-time signals at N sensors, or it could be the result of oversampling (fractional sampling) at fewer than N sensors [1]-[3].

In [4],[5] an iterative, inverse filter criteria based approach has been developed for deconvolution of multichannel non-Gaussian processes using the fourth-order normalized cumulants of the inverse filtered data at zero-lag. The approach is input-iterative, i.e., the inputs are extracted and removed one-by-one. The matrix impulse response is then obtained by cross-correlating the extracted inputs with the observed outputs. In this paper we develop a stochastic gradient-based "recursification" of all of the batch optimization steps in [4],[5]. An interesting input-iterative adaptive approach using prewhitened observations and the fourth-order cumulant of the inverse-filtered data at zero-lag has been considered in [11] and [11]. The inverse filter is constrained to have a lossless filter structure which is realized using a lossless lattice filter. Such a restriction can lead to ill-conditioning of the algorithm of [11] as one iteratively extracts input sequences. A fix to this is proposed in [12] but it works only for the two-input case. Refs. [11] and [12] are restricted to $M = N$ whereas in this paper we allow $N \geq M$, a common occurrence in array processing. Moreover, in this paper we perform no prewhitening, rather we operate directly on the given measurements.

This work was supported by the National Science Foundation under Grant MIP-9312559 and by the Office of Naval Research under Grant N00014-97-1-0822.

2. MODEL ASSUMPTIONS

The following assumptions are made concerning the system model (1-1) and (1-2):

(AS1) The vector sequence $\{\mathbf{w}(k)\}$ is zero-mean, temporally i.i.d. (independent and identically distributed) and spatially independent, i.e., various components of $\mathbf{w}(k)$ are independent of each other but not necessarily identically distributed. Assume that the fourth-order cumulant (see (3-1) later) of all the components of $\mathbf{w}(k)$ are nonzero but not necessarily negative.

(AS2) If it is an infinite impulse response (IIR) model, then (1-2) is assumed to be the result of a finite-dimensional multichannel ARMA model such that the model matrix impulse response function is exponentially stable, i.e., $\| [f_{ij}(l)] \| < a\beta^{|l|}$ for some $0 < a < \infty$ and $0 < \beta < 1$ where $[f_{ij}(l)]$ denotes a matrix with its ij-th element as $f_{ij}(l)$.

(AS3) $N \geq M$, i.e. at least as many outputs as inputs.

(AS4) $\text{Rank}\{\mathcal{F}(z)\} = M$ for any $|z| = 1$.

Notice that we allow the fourth-order cumulants of some components of $\mathbf{w}(k)$ to be positive. Moreover, we do not require $E\{w_j^2(k)\} = 0$ if the component $w_j(k)$ has negative fourth cumulant; this is in contrast to the CMA/Godard algorithm-based approaches where we also must have $E\{w_j^2(k)\} = 0$ in addition to negative fourth cumulant of $w_j(k)$. The objective is to recover $w_j(k)$ $\forall j$. It has been pointed out in [9] that for complex MIMO channel-equalizer cascades, but with real-valued sources, the CMA/Godard costs will have some undesirable global minima. "The real and imaginary parts of each equalizer output after convergence, may correspond to different user signals" [9]. It has been shown in [9] that the reason for this is that such real-valued signals are asymmetric (i.e. $E\{w_j^2(k)\} \neq 0$). Such a misconvergence can not occur for the cost function (3-3) considered in this paper [5].

3. AN ITERATIVE SOLUTION

In this section we briefly discuss the batch (non-recursive) approach of [4],[5]; its adaptive version is developed in Sec. 4. Let $\text{CUM}_4(w)$ denote the fourth-order cumulant of a complex-valued scalar zero-mean random variable w, defined as

$$\text{CUM}_4(w) := E\{|w|^4\} - 2[E\{|w|^2\}]^2 - |E\{w^2\}|^2. \quad (3-1)$$

Consider an $1 \times N$ row-vector polynomial equalizer (filter) $\mathcal{C}^T(z)$, with its j-th entry denoted by $\mathcal{C}_j(z)$, operating on the data vector $\mathbf{y}(k)$. Let the equalizer output be denoted by $e(k)$:

$$e(k) = \sum_{i=1}^{N} \mathcal{C}_i(z) y_i(k). \qquad (3-2)$$

Following [4] consider maximization of the cost

$$J := \frac{|\mathrm{CUM}_4(e(k))|}{[E\{|e(k)|^2\}]^2} \qquad (3-3)$$

for designing a linear equalizer to recover one of the inputs. It is shown [4] that when (3-3) is maximized w.r.t. $\mathcal{C}(z)$, then (3-2) reduces to

$$e(k) = d w_{j_0}(k - k_0), \qquad (3-4)$$

where d is some complex constant, k_0 is some integer, j_0 indexes some input out of the given M inputs, i.e., the equalizer output is a possibly scaled and shifted version of one of the system inputs. It has been established in [5] that under **(AS1)**-**(AS4)** and no noise, such a solution exists and if doubly-infinite equalizers are used, then all locally stable stationary points of the given cost w.r.t. the equalizer coefficients are also characterized by solutions such as (3-4).

A source-iterative solution is given by:

Step 1. Maximize (3-3) w.r.t. the equalizer $\mathcal{C}(z)$ to obtain (3-4).

Step 2. Cross-correlate $\{e(k)\}$ (of (3-4)) with the given data (2-2) and define a possibly scaled and shifted estimate of $f_{ij_0}(\tau)$ as

$$\widehat{f}_{ij_0}(\tau) := \frac{E\{y_i(k)e^*(k-\tau)\}}{E\{|e(k)|^2\}} \qquad (3-5)$$

where $F_{ij}(z) = \sum_{l=-\infty}^{\infty} f_{ij}(l) z^{-l}$. Consider now the reconstructed contribution of $e(k)$ to the data $y_i(k)$ $(i = 1, 2, \cdots, N)$, denoted by $\widehat{y}_{i,j_0}(k)$:

$$\widehat{y}_{i,j_0}(k) := \sum_l \widehat{f}_{ij_0}(l) e(k-l). \qquad (3-6)$$

Step 3. Remove the above contribution from the data to define the outputs of a MIMO system with N outputs and $M-1$ inputs. These are given by

$$y_i'(k) := y_i(k) - \widehat{y}_{i,j_0}(k). \qquad (3-7)$$

Step 4. If $M > 1$, set $M \leftarrow M-1$, $y_i(k) \leftarrow y_i'(k)$, and go back to **Step 1**, else quit.

In practice, all the expectations in (3-5) are replaced with their sample averages over appropriate data records.

It has been shown in [4],[5] that

$$\widehat{y}_{i,j_0}(k) = \sum_l f_{ij_0}(l) w_{j_0}(k-l), \qquad (3-8)$$

i.e., we have decomposed the observations at the various sensors into its independent components: $\widehat{y}_{i,j_0}(k)$ in (3-8) represents the contribution of $\{w_{j_0}(k)\}$ to the i-th sensor achieving **blind signal separation**.

Theorem 1[5]: Given the model (1-2) such that $\mathbf{n}(k) \equiv 0$ and given the true 4th-order and 2nd-order cumulant functions of the model output $\{\mathbf{y}(k)\}$ such that conditions **(AS1)**-**(AS4)** hold true. Suppose that doubly infinite equalizers are used in steps 1–4 of the iterative procedure of Sec. 3. Then this procedure yields a transfer function $\mathcal{A}(z)$ satisfying

$$\mathcal{A}(z) = \mathcal{F}(z)\mathbf{DAP} \quad \bullet \qquad (3-9)$$

The results of [4],[5] are based upon the use of doubly-infinite inverse filters. If we assume that $\mathcal{F}(z)$ has finite impulse response (FIR) and rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), then finite length inverse filters suffice. For an analysis and further elaborations, see [6] and [7] where a Godard cost function is considered but the results of [6] and [7] can be easily modified to apply to the cost (3-3). The following result follows from [5] and [7].

Theorem 2: Given the FIR model (1-2) such that $\mathbf{n}(k) \equiv 0$ and conditions **(AS1)** and **(AS4)** hold true. Suppose that steps 1–4 of the iterative procedure of Sec. 3 are used and the record length tends to infinity. Then this procedure yields a transfer function $\mathcal{A}(z)$ satisfying (3-9) if one of the following holds true:

(A) Rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), and doubly-infinite equalizers are used.

(B) Rank$\{\mathcal{F}(z)\} = M$ for any z (including $z = \infty$ but excluding $z = 0$), $\mathcal{F}(z)$ is column-reduced and FIR equalizers with length $L_e \geq (2M-1)L_c - 1$ are used where $L_c =$ channel length. $\bullet$

4. ADAPTIVE ALGORITHM

In this section we develop a stochastic gradient-based "recursification" of all of the batch optimization steps discussed in Sec. 3. Theorems 1 and 2 of Sec. 3 motivate and justify the algorithm developed in this section.

4.1. First Stage Maximization of Normalized Fourth Cumulant

Let the length of the equalizer $\mathcal{C}(z)$ be L_e and let

$$\mathcal{C}_i(z) = \sum_{l=0}^{L_e-1} c_i(l) z^{-l}. \qquad (4-1)$$

This allows us to rewrite (3-2) as

$$e(k) = \sum_{i=1}^{N} \sum_{l=0}^{L_e-1} c_i(l) y_i(k-l) = \mathbf{C}^T \mathbf{Y}(k) \qquad (4-2)$$

where

$$\mathbf{Y}(k) = [\ Y_1^T(k) \quad Y_2^T(k) \quad \cdots \quad Y_N^T(k)\ ]^T, \qquad (4-3)$$

$$Y_i(k) = [\ y_i(k) \quad y_i(k-1) \quad \cdots \quad y_i(k-L_e+1)\ ]^T, \qquad (4-4)$$

$$\mathbf{C}(k) = [\ C_1^T \quad C_2^T \quad \cdots \quad C_N^T\ ]^T, \qquad (4-5)$$

and

$$C_i = [\ c_i(0) \quad c_i(1) \quad \cdots \quad c_i(L_e-1)\ ]^T. \qquad (4-6)$$

Define

$$m_4 = E\{|e(k)|^4\}, \quad m_2 = E\{|e(k)|^2\}, \quad \widetilde{m}_2 = E\{e^2(k)\}. \qquad (4-7)$$

Then showing explicit dependence upon $\mathbf{C}$, (3-3) may be rewritten as

$$J(\mathbf{C}) = \mathrm{sgn}(\gamma_4) \left[\frac{m_4 - |\widetilde{m}_2|^2}{m_2^2} - 2 \right] \qquad (4-8)$$

where

$$\gamma_4 = m_4 - 2\,m_2^2 - |\widetilde{m}_2|^2. \qquad (4-9)$$

Let $\nabla_{\mathbf{C}}$ denote a gradient operator (w.r.t. a vector $\mathbf{C}$). We will follow [7] in formally defining the complex derivatives. Then we have

$$\nabla_{\mathbf{C}^*}\, e(k) = 0 \quad \text{and} \quad \nabla_{\mathbf{C}^*}\, e^*(k) = \mathbf{Y}^*(k). \qquad (4-10)$$

Using the above results in (4-7) we have

$$\nabla_{\mathbf{C}^*} m_4 = 2E\{e^2(k)e^*(k)\mathbf{Y}^*(k)\},$$

$$\nabla_{\mathbf{C}^*}\, m_2 = E\{e(k)\mathbf{Y}^*(k)\} \qquad (4-11)$$

and

$$\nabla_{\mathbf{C}^*}\, \widetilde{m}_2 = 0, \quad \nabla_{\mathbf{C}^*} \widetilde{m}_2^* = 2\,E\{e^*(k)\mathbf{Y}^*(k)\}. \quad (4-12)$$

Using (4-8)-(4-12) and after some simplification, we have

$$\nabla_{\mathbf{C}^*} J(\mathbf{C}) = \frac{2\,\mathrm{sgn}(\gamma_4)}{m_2^3}\left\{m_2 E\{|e(k)|^2 e(k)\mathbf{Y}^*(k)\}\right.$$

$$\left. -\widetilde{m}_2 m_2 E\{e^*(k)\mathbf{Y}^*(k)\} - [m_4 - |\widetilde{m}_2|^2]E\{e(k)\mathbf{Y}^*(k)\}\right\}. \qquad (4-13)$$

We will use a stochastic gradient method for recursification of maximization of $J(\mathbf{C})$ using an 'instantaneous' gradient as an estimate of (4-13). Given the estimate $\mathbf{C}(k-1)$ of the tap-gains at time $k-1$, the stochastic gradient method computes the update $\mathbf{C}(k)$ at time k as

$$\widetilde{\mathbf{C}}(k) = \mathbf{C}(k-1) + \mu_1\, \nabla_{\mathbf{C}^*}\, J_k(\mathbf{C}(k-1)) \qquad (4-14)$$

$$\mathbf{C}(k) = \frac{\widetilde{\mathbf{C}}(k)}{\|\widetilde{\mathbf{C}}(k)\|} \qquad (4-15)$$

where μ_1 is the update step-size and $\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k-1))$ is an instantaneous gradient of the cost J (w.r.t. $\mathbf{C}^*$) at time k evaluated at $\mathbf{C}(k-1)$. Since the cost J is invariant to any scaling of $\mathbf{C}$, we normalize $\mathbf{C}$ in (4-15) to have a unit norm. From (4-13) we have the approximation

$$\nabla_{\mathbf{C}^*} J_k(\mathbf{C}(k)) =$$

$$\mathrm{sgn}(\gamma_{4k})\frac{2}{m_{2k}^3}\left\{\left[m_{2k}\left(e^2(k) - \widetilde{m}_{2k}\right)e^*(k)\right.\right.$$

$$\left.\left. - \left(m_{4k} - |\widetilde{m}_{2k}|^2\right)e(k)\right]\mathbf{Y}^*(k)\right\} \qquad (4-16)$$

where

$$m_{2k} = (1-\mu_2)m_{2(k-1)} + \mu_2|e(k)|^2, \qquad (4-17)$$

$$\widetilde{m}_{2k} = (1-\mu_2)\widetilde{m}_{2(k-1)} + \mu_2 e^2(k), \qquad (4-18)$$

$$m_{4k} = (1-\mu_2)m_{4(k-1)} + \mu_2|e(k)|^4, \qquad (4-19)$$

$$\gamma_{4k} = m_{4k} - 2\,m_{2k}^2 - |\widetilde{m}_{2k}|^2 \qquad (4-20)$$

and

$$e(k) = \mathbf{C}^T(k)\mathbf{Y}(k). \qquad (4-21)$$

In (4-17)-(4-19) the various quantities represent estimates based upon sample averaging, the (exponential window) memory being controlled by the forgetting factor μ_2 ($0 < \mu_2 < 1$). The initializations for (4-17)-(4-19) are: $m_{20} = m_{40} = \widetilde{m}_{20} = 0$.

4.2. First Stage Signal Cancellation

Now we discuss implementation of (3-6) via sample averaging using an exponential window controlled by a forgetting factor μ_3. Define ($L_1, L_2 > 0$)

$$\mathbf{E}(k) = [\ e(k+L_1) \quad e(k+L_1-1) \quad \cdots \quad e(k-L_2)\ ]^T \qquad (4-22)$$

and

$$\mathbf{F}_i = [\ \widetilde{f}_i(-L_1) \quad \widetilde{f}_i(-L_1+1) \quad \cdots \quad \widetilde{f}_i(L_2)\ ]^T. \qquad (4-23)$$

By Sec. 3, when (4-8) is maximized, $e(k)$ satisfies (3-4) so that for suitable choice of L_1 and L_2, there exists a $j_0 \in \{1, 2, \cdots, M\}$ such that

$$\sum_l f_{ij_0}(l)w_{j_0}(k-l) = \mathbf{F}_i^T\mathbf{E}(k), \quad i = 1, 2, \cdots, N. \qquad (4-24)$$

In order to implement (3-6) and (3-7), we need recursive estimates of $\mathbf{F}_i$. The estimate $\mathbf{F}_i(k)$ of $\mathbf{F}_i$ at time k is provided by

$$\mathbf{F}_i(k) = \mathbf{R}_i(k)/m_{ee}(k) \qquad (4-25)$$

where

$$m_{ee}(k) = (1-\mu_3)m_{ee}(k-1) + \mu_3|e(k)|^2, \qquad (4-26)$$

$$\mathbf{R}_i(k) = (1-\mu_3)\mathbf{R}_i(k-1) + \mu_3 y_i(k)\mathbf{E}^*(k). \qquad (4-27)$$

4.3. Multistage Algorithm

In Secs. 4.1 and 4.2 we discussed the first stage of the algorithm where we have N sensors and M sources. Now we put it all together following the source-iterative solution of Sec. 3 and discuss extraction of M sources including the cancellation of the extracted sources. We will use the superscript (m) to denote the various quantities pertaining to stage m. These have been used previously in Secs. 4.1 and 4.2 without this superscript; e.g. $\mathbf{C}^{(m)}(k)$ now denotes the estimate of the tap-gain vector at time k at stage m, etc.

Initialization:

$$\mathbf{Y}^{(1)}(k) = \quad \text{as in (4-3)} \qquad (4-28)$$

DO FOR $m = 1, 2, \cdots, M$:

$$\widetilde{\mathbf{C}}^{(m)}(k) = \mathbf{C}^{(m)}(k-1) + \mu_1\, \nabla_{\mathbf{C}^*}\, J_k^{(m)}(\mathbf{C}^{(m)}(k-1)) \qquad (4-29)$$

$$\mathbf{C}^{(m)}(k) = \frac{\widetilde{\mathbf{C}}^{(m)}(k)}{\|\widetilde{\mathbf{C}}^{(m)}(k)\|} \qquad (4-30)$$

where

$$\nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k)) =$$

$$\mathrm{sgn}(\gamma_{4k}^{(m)})\frac{2}{m_{2k}^{(m)3}}\left\{\left[m_{2k}^{(m)}\left(e^{(m)2}(k) - \widetilde{m}_{2k}^{(m)}\right)e^{(m)*}(k)\right.\right.$$

$$\left.\left. - \left(m_{4k}^{(m)} - |\widetilde{m}_{2k}^{(m)}|^2\right)e^{(m)}(k)\right]\mathbf{Y}^{(m)*}(k)\right\}, \qquad (4-31)$$

$$m_{2k}^{(m)} = (1-\mu_2)m_{2(k-1)}^{(m)} + \mu_2|e^{(m)}(k)|^2, \qquad (4-32)$$

$$\widetilde{m}_{2k}^{(m)} = (1-\mu_2)\widetilde{m}_{2(k-1)}^{(m)} + \mu_2 e^{(m)2}(k), \qquad (4-33)$$

$$m_{4k}^{(m)} = (1-\mu_2)m_{4(k-1)}^{(m)} + \mu_2|e^{(m)}(k)|^4, \qquad (4-34)$$

$$\gamma_{4k}^{(m)} = m_{4k}^{(m)} - 2\,m_{2k}^{(m)2} - |\widetilde{m}_{2k}^{(m)}|^2 \qquad (4-35)$$

and

$$e^{(m)}(k) = \mathbf{C}^{(m)T}(k)\mathbf{Y}^{(m)}(k). \qquad (4-36)$$

Set

$$\widehat{y}_i^{(m)}(k) = \mathbf{F}_i^{(m)T}(k)\mathbf{E}^{(m)}(k) \qquad (4-37)$$

where $\widehat{y}_i^{(m)}(k)$ represents (cf. (3-7)) the contribution of the extracted source at the $m-$th stage to the measurement at time k at the $i-$th sensor, and where

$$\mathbf{F}_i^{(m)}(k) = \mathbf{R}_i^{(m)}(k)/m_{ee}^{(m)}(k), \qquad (4-38)$$

$$m_{ee}^{(m)}(k) = (1-\mu_3)m_{ee}^{(m)}(k-1) + \mu_3|e^{(m)}(k)|^2, \qquad (4-39)$$

$$\mathbf{E}^{(m)}(k) = \left[\begin{array}{ccc} e^{(m)}(k+L_1) & \cdots & e^{(m)}(k-L_2) \end{array}\right]^T \qquad (4-40)$$

and

$$\mathbf{R}_i^{(m)}(k) = (1-\mu_3)\mathbf{R}_i^{(m)}(k-1) + \mu_3 y_i^{(m)}(k)\mathbf{E}^{(m)*}(k). \qquad (4-41)$$

Define

$$\widehat{Y}_i^{(m)}(k) = \left[\begin{array}{ccc} \widehat{y}_i^{(m)}(k) & \cdots & \widehat{y}_i^{(m)}(k-L_e+1) \end{array}\right]^T, \qquad (4-42)$$

Set

$$\mathbf{Y}^{(m+1)}(k) = \left[\begin{array}{ccc} Y_1^{(m+1)T}(k) & \cdots & Y_N^{(m+1)T}(k) \end{array}\right]^T \qquad (4-43)$$

where

$$Y_i^{(m+1)}(k) = Y_i^{(m)}(k) - \widehat{Y}_i^{(m)}(k). \qquad (4-44)$$

ENDDO

The sequence $\{e^{(m)}(k)\}$ in (4-36) represents the equalized (up to a scale factor and time delay) source at stage m.

Running Cost. To monitor the convergence of the equalizers in various stages of the algorithm, it is useful to calculate a running cost (4-8) without the sign. Let $J_k^{(m)}$ denote the running cost for the $m-$th stage at time k, given by

$$J_k^{(m)} = \frac{m_{4k}^{(m)} - |\widetilde{m}_{2k}^{(m)}|^2}{m_{2k}^{(m)2}} - 2 \qquad (4-45)$$

where

$$m_{2k}^{(m)} = (1-\mu_4)m_{2(k-1)}^{(m)} + \mu_4|e^{(m)}(k)|^2, \qquad (4-46)$$

$$\widetilde{m}_{2k}^{(m)} = (1-\mu_4)\widetilde{m}_{2(k-1)}^{(m)} + \mu_4 e^{(m)2}(k), \qquad (4-47)$$

$$m_{4k}^{(m)} = (1-\mu_4)m_{4(k-1)}^{(m)} + \mu_4|e^{(m)}(k)|^4. \qquad (4-48)$$

For all of the simulations presented in Sec. 5, we took $\mu_4 = 0.002$.

5. SIMULATION EXAMPLES

5.1. Example 1: 3 BPSK Sources and 7 Sensors

We consider a wireless communications scenario with three ($M = 3$) BPSK user signals arriving at a uniform linear array of $N = 7$ sensors via a frequency selective multipath channel. The array elements are spaced half a wavelength apart. The array measurements are assumed to be sampled at baud rate (for simulation convenience only) with sampling interval T seconds and the three sources have the same baud rate. The relative time delay τ (relative to the first arrival), the angle of arrival θ (in degrees w.r.t. the array broadside) and the relative attenuation factor (amplitude) α for various sources were selected as:

$$w_1: (\tau,\theta,\alpha) = (0T, 10^\circ, 0.5),\ (1T, 50^\circ, 0.75)$$

$$w_2: (\tau,\theta,\alpha) = (0T, -20^\circ, 0.5),\ (1T, 45^\circ, 0.45),\ (2T, 15^\circ, -0.65)$$

$$w_3: (\tau,\theta,\alpha) = (0T, -35^\circ, 0.7),\ (1T, -5^\circ, 0.4).$$

Thus the signals w_1 and w_3 propagate through two paths whereas w_2 passes through three paths. The signals arriving at the array were normalized such that the signal powers for users 1 and 2 are equal, and 3dB higher than the signal power for user 3. Additive white (both temporally and spatially) Gaussian noise was added to the array measurements to achieve a signal-to-noise-ratio (SNR) of 11.55dB (ratio = 100/7) for the strongest user(s).

The proposed approach was applied with $M = 3$ equalizers and $M - 1 = 2$ signal cancellers running in parallel, each successive equalizer put in operation after waiting for 200 samples (symbols) w.r.t. the previous stage. The equalizer length was chosen to be 5 taps per sensor ($L_e = 5$ in (4-2)). The initial guess for the tap gains was taken to be center-tap initialization: set $c_i(2) = 1$ for $i = m$ for the $m-$th stage equalizer ($m = 1, 2, 3$) with the remaining tap gains set to zero. The algorithm step sizes and forgetting factors for each stage m were chosen as: $\mu_1 = 0.003$ in (4-29), $\mu_2 = 0.015$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41). For the running cost (4-45) computation we selected $\mu_4 = 0.002$ in (4-46)-(4-48). The parameters L_1 and L_2 in (4-40) (see also (4-22) and (4-23)) were selected as $L_1 = 15$ and $L_2 = 6$.

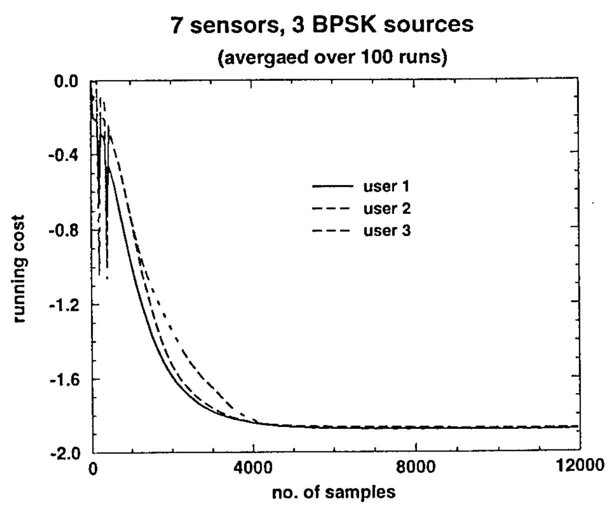


Fig. 1. Average running cost for Example 1.

Fig. 1 shows the evolution of the average running cost $J_k^{(m)}$ (see (4-45)), averaged over 100 Monte Carlo runs after 'assigning' each equalizer cost to its corresponding extracted source. For BPSK sources the 4th-order normalized cumulant equals -2; therefore, at convergence, the running cost (4-45) should be close to -2. In Fig. 1 we see these values to be around -1.89 which is largely a consequence of noise in the data which affects only the denominator of (4-45) making it larger than it should be. Table 1 shows the signal-to-interference-and-noise ratio (SINR) and the probability of error P_e at the output of each equalizer at selected time instants, averaged over 100 Monte Carlo runs and 3000 symbols. [The equalizer tap gains at the chosen time instants were 'frozen' and used to equalize data of length 3000 symbols in order to calculate SINR and P_e. The equalized data were rotated, scaled and shifted before calculating the two performance measures.] It is seen from Fig. 1 and Table 1 that the proposed approach works well.

5.2. Example 2: 2 Mixed Sources and 5 Sensors

In this example we consider a 4-QAM user signal w_1 (4th normalized cumulant as -1) and a non-communications signal w_2 consisting of an i.i.d. complex Gaussian-mixture (independent and identically distributed real and imaginary parts with the real part being $\mathcal{N}(0,1)$ with probability 0.9 and $\mathcal{N}(0,4)$ with probability 0.1) with 4th normalized cumulant as 0.7433 . The multipath channels for the two signals were selected as:

$$w_1: \ (\tau,\theta,\alpha) = (0T, 10^\circ, 0.5), \ (1T, 50^\circ, 0.75)$$

$$w_2: \ (\tau,\theta,\alpha) = (0T, -20^\circ, 0.5), \ (1T, 45^\circ, 0.45), \ (2T, 15^\circ, -0.65).$$

The two signals have equal power. Additive white Gaussian noise was added to the array measurements to achieve an SNR of 13dB (ratio = 20) for each user signal.

The proposed approach was applied with $M = 2$ equalizers and $M - 1 = 1$ signal cancellers running in parallel, the second equalizer put in operation after waiting for 200 samples. The equalizer length was chosen to be 5 taps per sensor ($L_e = 5$ in (4-2)). The initial guess for the tap gains was taken to be center-tap initialization: set $c_i(2) = 1$ for $i = m$ for the $m-$th stage equalizer ($m = 1,2$) with the remaining tap gains set to zero. The algorithm step sizes and forgetting factors were chosen as: $\mu_1 = 0.0005$ in (4-29), $\mu_2 = 0.015$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41) when $\gamma_{4k}^{(m)} \leq 0$ (see (4-35)), and $\mu_1 = 0.0001$ in (4-29), $\mu_2 = 0.003$ in (4-32)-(4-34) and $\mu_3 = 0.0005$ in (4-39) and (4-41) when $\gamma_{4k}^{(m)} > 0$. The parameters L_1 and L_2 in (4-40) were selected as $L_1 = 15$ and $L_2 = 6$. For the running cost (4-45) computation we selected $\mu_4 = 0.002$ in (4-46)-(4-48). The counterparts to Fig. 1 and Table 1 are now shown in Fig. 2 and Table 2, respectively.

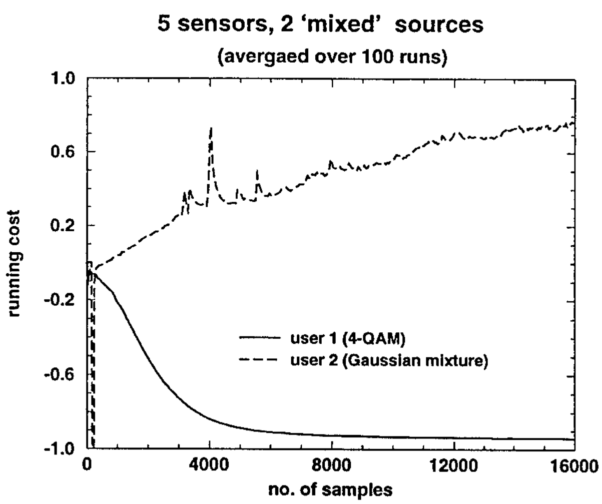


Fig. 2. Average running cost for Example 2.

6. REFERENCES

[1] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channels," *IEEE Trans. Information Theory*, vol. IT-40, pp. 340-349, March 1994.

[2] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 516-525, Feb. 1995.

[3] D.T.M. Slock, "Blind joint equalization of multiple synchronous mobile users using oversampling," in *Proc. 28th Asilomar Conf. Signals, Systems, Computers*, pp. 1154-1158, Pacific Grove, CA, Oct.-Nov. 1994.

[4] J.K. Tugnait, "Identification of multichannel linear non-Gaussian processes using higher-order statistics," in *Proc. 29th Annual Asilomar Conf. Signals Sys. Comp.*, Pacific Grove, CA, Oct. 29 - Nov. 1, 1995.

[5] J.K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 658-672, March 1997.

[6] J.K. Tugnait, "Spatio-temporal signal processing for blind separation of multichannel signals," in *Digital Signal Processing Technology*, Joseph Picone, Editor, *Proc. SPIE 2750*, pp. 88-103, 1996.

[7] J.K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Transactions on Signal Processing*, vol. SP-45, pp. 268-271, Jan. 1997.

[8] D.H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. IEE*, vol. 130, pts. F & H, pp. 11-16, Feb. 1983.

[9] C.B. Papadias, "On the existence of undesirable global minima of Godard equalizers," in *Proc. IEEE 1997 Intern. Conf. Acoustics, Speech, Signal Processing*, pp. V-3941-3944, April 24-27, 1997, Munich, Germany.

[10] Special Issue, *IEEE Transactions on Signal Processing*, vol. SP-45, Jan. 1997.

[11] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals using adaptive FIR lossless filters," in *Proc. EUSIPCO 92*, pp. 1061-1064, Brussels, Belgium, Aug. 1992.

[12] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. ICC 93*, pp. 1160-1164, Geneva, Switzerland, June 1993.

TABLE 1

# of samples	User 1 SINR(dB)	P_e
4000	14.89	$< 3 \times 10^{-4}$
6000	15.03	$< 3 \times 10^{-4}$
8000	15.13	$< 3 \times 10^{-4}$
12000	15.14	$< 3 \times 10^{-4}$
	User 2	
4000	14.25	0.0099
6000	14.81	$< 3 \times 10^{-4}$
8000	14.94	$< 3 \times 10^{-4}$
12000	14.97	$< 3 \times 10^{-4}$
	User 3	
4000	14.84	$< 3 \times 10^{-4}$
6000	15.13	$< 3 \times 10^{-4}$
8000	15.19	$< 3 \times 10^{-4}$
12000	15.23	$< 3 \times 10^{-4}$

TABLE 2

# of samples	User 1 SINR(dB)	User 1 P_e	User 2 SINR(dB)
8000	14.12	0.0072	5.52
12000	14.79	$< 3 \times 10^{-4}$	9.28
16000	15.06	$< 3 \times 10^{-4}$	11.49

[8]

BLIND EQUALIZATION OF I.I.R. SINGLE-INPUT MULTIPLE-OUTPUT CHANNELS WITH COMMON ZEROS USING SECOND-ORDER STATISTICS

Bin Huang *Jitendra K. Tugnait*

Department of Electrical Engineering
Auburn University, Auburn, Alabama 36849, USA
E-mail: huangbi@eng.auburn.edu tugnait@eng.auburn.edu

ABSTRACT

The problem of blind equalization of SIMO (single-input multiple-output) communications channels is considered using only the second-order statistics of the data. Such models arise when a single receiver data is fractionally sampled (assuming that there is excess bandwidth), or when an antenna array is used with or without fractional sampling. We focus on direct design of finite-length MMSE (minimum mean-square error) blind equalizers. Unlike the past work on this problem, we allow infinite impulse response (IIR) channels. Our approaches also work when the "subchannel" transfer functions have common zeros so long as the common zeros are minimum-phase zeros. Illustrative simulation examples are provided.

1. INTRODUCTION

Consider a discrete-time SIMO system with N outputs and one input. The i-th component of the output at time k is given by

$$y_i(k) = \mathcal{F}_i(z)w(k) + n_i(k), \quad i = 1, 2, \cdots, N, \qquad (1-1)$$

$$\Longrightarrow \quad \mathbf{y}(k) = \mathcal{F}(z)w(k) + \mathbf{n}(k) = \mathbf{s}(k) + \mathbf{n}(k), \quad (1-2)$$

where $\mathbf{y}(k) = [y_1(k) \vdots y_2(k) \vdots \cdots \vdots y_N(k)]^T$, similarly for $\mathbf{s}(k)$ and $\mathbf{n}(k)$, and z is the $\mathcal{Z}-$transform variable as well as the backward-shift operator (i.e., $z^{-1}w(k) = w(k-1)$, etc.). The sequence $w(k)$ is the (single) input at sampling time k, $y_i(k)$ is the i-th noisy output, $s_i(k)$ is the i-th noise-free output, $n_i(k)$ is the additive measurement noise,

$$\mathcal{F}(z) := \sum_{l=0}^{\infty} \mathbf{F}_l z^{-l} \qquad (1-3)$$

and $\mathcal{F}_i(z) = \sum_{l=0}^{\infty} f_i(l) z^{-l}$ is the scalar transfer function with $w(k)$ as the input and $y_i(k)$ as the output; it represents the $i-$th subchannel. We allow all of the above variables to be complex-valued.

Such models arise in several useful baseband-equivalent digital communications and other applications. A case of some interest is that of fractionally-spaced samples of a single baseband received signal leading to a SIMO model [1],[4],[8]. Alternatively, a similar model can be derived when we have a single signal impinging upon an antenna array with N elements [5]. A similar model arises if we have an antenna array coupled with fractional sampling at each array-element [5]. In these applications one of the objectives is to recover the inputs $w(k)$ given the noisy measurements but not given the knowledge of the system transfer function. An overwhelming number of papers (see [4],[5],[9]-[12] and references therein) have concentrated on a two-step procedure: first estimate the channel impulse response (IR) and then design an equalizer using the estimated channel. A fundamental restriction in these works is that the channel is FIR with no common zeros among the various subchannels. A few (see [1]and [13], e.g.) have proposed direct design of the equalizer bypassing channel estimation. Still they assume FIR channels with no common zeros.

In this paper we allow IIR channels. We will also allow common zeros so long as they are minimum-phase. Finally, in the presence of nonminimum-phase common zeros, our proposed approach equalizes the spectrally-equivalent minimum-phase counterpart of $\mathcal{F}(z)$; it does not "fall apart" unlike quite a few existing approaches. We should note that our proposed approach is inspired by [1]. Unlike [1] our approach applies to antenna arrays since we do not require that $f_1(0) \neq 0$ but $f_i(0) = 0$ for $i = 2, 3, \cdots, N$ (as in [1]).

This work was supported by NSF Grant MIP-9312559 and by ONR Grant N00014-97-1-0822.

2. PRELIMINARIES

2.1. FIR Inverses

Let $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ where $\mathcal{A}(z) = 1 + \sum_{i=1}^{n_a} a_i z^{-i}$ is 1×1 and $\mathcal{B}(z) = \sum_{i=0}^{n_b} \mathbf{B}_i z^{-i}$ is $N \times 1$. Assume

(H1) $N > 1$.
(H2) Rank$\{\mathcal{B}(z)\} = 1\ \forall z$ including $z = \infty$ but excluding $z = 0$, i.e., $\mathcal{B}(z)$ is irreducible [7, Sec. 6.3].
(H3) $\mathcal{A}(z) \neq 0$ for $|z| \geq 1$.

It has been shown in [6] (using some results from [2]) that under **(H1)-(H3)** there exists a finite degree left-inverse (not necessarily unique) of $\mathcal{F}(z)$:

$$\mathcal{G}(z)\mathcal{F}(z) = 1 \qquad (2-1)$$

where $\mathcal{G}(z)$ is $1 \times N$ given by

$$\mathcal{G}(z) = \sum_{l=0}^{L_e} \mathbf{G}_l z^{-l} \quad \text{for any} \quad L_e \geq n_a + n_b - 1. \quad (2-2)$$

Remark 1: The left-inverse $\mathcal{G}(z)$ of $\mathcal{F}(z)$ consists of two parts: $\mathcal{G}(z) = \mathcal{G}_B(z)\mathcal{A}(z)$ where $\mathcal{G}_B(z)\mathcal{B}(z) = 1$ so that $\mathcal{G}(z)\mathcal{F}(z) = \mathcal{G}_B(z)\mathcal{A}(z)\mathcal{A}^{-1}(z)\mathcal{B}(z) = \mathcal{G}_B(z)\mathcal{B}(z) = 1$. Finite length left-inverses of FIR SIMO channels have been subject of intense research activities [4]-[6],[8]-[13].

2.2. Linear Innovations Representations

Assume further the following:

(H4) $\{w(k)\}$ is zero-mean, white. Take $E\{|w(k)|^2\} = 1$.

Lemma 1. Under **(H1)-(H4)**, $\{\mathbf{s}(k)\}$ may be represented as

$$\mathbf{s}(k) = -\sum_{i=1}^{M} \mathbf{D}_i \mathbf{s}(k-i) + I_s(k) \qquad (2-3)$$

where $M = n_a + n_b - 1$, $\mathbf{D}_i$'s are some $N \times N$ matrices such that $\det(\mathcal{D}(z)) \neq 0$ for $|z| \geq 1$, $\mathcal{D}(z) = I + \sum_{i=1}^{M} \mathbf{D}_i z^{-i}$ and $\{I_s(k)\}$ is a zero-mean white $N \times 1$ random sequence (linear innovations for $\{\mathbf{s}(k)\}$) with

$$E\{I_s(k)I_s^{\mathcal{H}}(k)\} = \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} \quad \text{and} \quad \|\mathbf{F}_0\|^{-2}\mathbf{F}_0^{\mathcal{H}} I_s(k) = w(k). \quad (2-4)$$

Proof: Consider the process

$$\mathbf{s}'(k) := \mathcal{A}(z)\mathbf{s}(k) = \mathcal{B}(z)w(k). \qquad (2-5)$$

By [9] and [14], under **(H1)**, **(H2)** and **(H4)**, we have

$$\mathbf{s}'(k) = -\sum_{i=1}^{n_b-1} \mathbf{D}_i'\mathbf{s}'(k-i) + I_s'(k) \qquad (2-6)$$

where $\mathbf{D}_i'$s are some $N \times N$ matrices such that $\det(\mathcal{D}'(z)) \neq 0$ for $|z| \geq 1$, $\mathcal{D}'(z) = I + \sum_{i=1}^{M} \mathbf{D}_i' z^{-i}$ and $\{I_s'(k)\}$ is a zero-mean white $N \times 1$ random sequence with

$$E\{I_s'(k)I_s'^{\mathcal{H}}(k)\} = \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} \text{ and } \|\mathbf{F}_0\|^{-2}\mathbf{F}_0^{\mathcal{H}} I_s'(k) = w(k). \qquad (2-7)$$

Since $\mathbf{s}(k) = \mathcal{A}^{-1}(z)\mathbf{s}'(k)$, it follows from (2-6) that (2-3) holds true with $I_s(k) \equiv I_s'(k)$ such that $\mathcal{D}(z) = \mathcal{A}(z)\mathcal{D}'(z)$. This completes the proof. □

Lemma 2. Let $\mathcal{R}_{ssL_e}$ denote a $[N(L_e+1)] \times [N(L_e+1)]$ matrix with its ij-th block element as $\mathbf{R}_{ss}(j-i) = E\{\mathbf{s}(k+j-i)\mathbf{s}^{\mathcal{H}}(k)\}$. Then under **(H1)**-**(H4)**, $\rho(\mathcal{R}_{ssL_e}) \leq NL_e + 1$ for $L_e \geq n_a + n_b - 1$ where $\rho(A)$ denotes the rank of A. •

Sketch of proof: It follows from Lemma 1 and (2-3) that

$$\begin{aligned} &[\ I \quad \mathbf{D}_1 \quad \cdots \quad \mathbf{D}_{n_a+n_b-1} \quad 0 \quad \cdots \quad 0\ ]\mathcal{R}_{ssL_e} \\ &= [\ \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} \quad 0 \quad \cdots \quad 0\ ]. \qquad (2-8) \end{aligned}$$

Apply Sylvester's inequality [7, p. 655] to (2-8) to deduce the desired result. □

3. BLIND EQUALIZATION: NO COMMON ZEROS

Assume that **(H1)**-**(H4)** hold true. In addition assume the following regarding the measurement noise:

(H5) $\{\mathbf{n}(k)\}$ is zero-mean with $E\{\mathbf{n}(k+\tau)\mathbf{n}^{\mathcal{H}}(k)\} = \sigma_n^2 I_{N\times N}$ where $I_{N\times N}$ is the $N \times N$ identity matrix.

3.1. Zero-Delay Zero-Forcing Equalizer

Using (1-3), (2-1) and (2-2), we have

$$\sum_{l=0}^{\infty} \mathbf{G}_{m-l}\mathbf{F}_l = \begin{cases} 1, & m = 0 \\ 0, & m = 1, 2, \cdots, \end{cases} \qquad (3-1)$$

leading to

$$[\ \mathbf{G}_0 \quad \mathbf{G}_1 \quad \cdots \quad \mathbf{G}_{L_e}\ ]\overline{\mathcal{S}} = [\ 1 \quad 0 \quad \cdots \quad \cdots\ ] \qquad (3-2)$$

where $\overline{\mathcal{S}}$ is the $(N(L_e+1)) \times \infty$ matrix given by

$$\overline{\mathcal{S}} = \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3 & \cdots & \cdots & \cdots \\ 0 & \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & & & \\ 0 & 0 & \cdots & 0 & \mathbf{F}_0 & \mathbf{F}_1 & \cdots \end{bmatrix}. \qquad (3-3)$$

Let $\overline{\mathcal{S}}^{\#}$ denote the pseudoinverse of $\overline{\mathcal{S}}$. By [15, Prop. 1], $\overline{\mathcal{S}}^{\#} = \overline{\mathcal{S}}^{\mathcal{H}}(\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}})^{\#}$. Then the minimum norm solution to the FIR equalizer is given by [15, Sec. 6.11]

$$[\ \mathbf{G}_0 \quad \mathbf{G}_1 \quad \cdots \quad \mathbf{G}_{L_e}\ ] = [\ \mathbf{F}_0^{\mathcal{H}} \quad 0 \quad \cdots \quad 0\ ](\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}})^{\#}. \qquad (3-4)$$

In a fashion similar to $\mathcal{R}_{ssL_e}$ in Lemma 2, let $\mathcal{R}_{yyL_e}$ denote a $[N(L_e+1)] \times [N(L_e+1)]$ matrix with its ij-th block element as $\mathbf{R}_{yy}(j-i) = E\{\mathbf{y}(k+j-i)\mathbf{y}^{\mathcal{H}}(k)\}$; define similarly $\mathcal{R}_{nnL_e}$ pertaining to the additive noise. Carry out an eigendecomposition of $\mathcal{R}_{yyL_e}$. Then the smallest $N-1$ eigenvalues of $\mathcal{R}_{yyL_e}$ equal σ_n^2 because under **(H1)**-**(H4)**, $\rho(\mathcal{R}_{ssL_e}) \leq NL_e + 1$ whereas $\rho(\mathcal{R}_{nnL_e}) = NL_e + N = \rho(\mathcal{R}_{yyL_e})$. Thus a consistent estimate $\widehat{\sigma}_n^2$ of σ_n^2 is obtained by taking it as the average of the smallest $N-1$ eigenvalues of $\widehat{\mathcal{R}}_{yyL_e}$, the data-based consistent estimate of $\mathcal{R}_{yyL_e}$.

Under (**H4**) and (**H5**),

$$(\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}}) = \mathcal{R}_{ssL_e} = \mathcal{R}_{yyL_e} - \mathcal{R}_{nnL_e} = \mathcal{R}_{yyL_e} - \sigma_n^2 I. \qquad (3-5)$$

Thus, $(\overline{\mathcal{S}}\,\overline{\mathcal{S}}^{\mathcal{H}})$ can be estimated from noisy data. However, we don't know $\mathbf{F}_0$. To this end, we seek an $N \times N$ FIR filter $\mathcal{G}_a(z) := \sum_{i=0}^{L_e} \mathbf{G}_{ai} z^{-i}$ satisfying

$$[\ \mathbf{G}_{a0} \quad \mathbf{G}_{a1} \quad \cdots \quad \mathbf{G}_{aL_e}\ ] = [\ I_{N\times N} \quad 0 \quad \cdots \quad 0\ ]\mathcal{R}_{ssL_e}^{\#}. \qquad (3-6)$$

Comparing (3-4) and (3-6) it follows that

$$[\ \mathbf{G}_0 \quad \mathbf{G}_1 \quad \cdots \quad \mathbf{G}_{L_e}\ ] = \mathbf{F}_0^{\mathcal{H}}[\ \mathbf{G}_{a0} \quad \mathbf{G}_{a1} \quad \cdots \quad \mathbf{G}_{aL_e}\ ] \qquad (3-7)$$

leading to

$$\sum_{i=0}^{L_e} \mathbf{G}_i z^{-i} =: \mathcal{G}(z) = \mathbf{F}_0^{\mathcal{H}}\mathcal{G}_a(z). \qquad (3-8)$$

In practice, therefore, we apply $\mathcal{G}_a(z)$ to the data leading to

$$\mathbf{v}(k) := \mathcal{G}_a(z)\mathbf{y}(k) = \mathbf{v}_s(k) + \mathcal{G}_a(z)\mathbf{n}(k) \qquad (3-9)$$

such that

$$\mathbf{F}_0^{\mathcal{H}}\mathbf{v}_s(k) = w(k) \qquad (3-10)$$

where

$$\mathbf{v}_s(k) := \mathcal{G}_a(z)[\mathbf{y}(k) - \mathbf{n}(k)] = \mathcal{G}_a(z)\mathbf{s}(k). \qquad (3-11)$$

In (3-10) $\{w(k)\}$ is a white scalar sequence (by assumption (**H4**)), however, $\{\mathbf{v}_s(k)\}$ is not necessarily a white vector sequence. Given the second-order statistics of $\{\mathbf{v}_s(k)\}$, how does one estimate $\mathbf{F}_0$ so that $\{w(k)\}$ satisfying (**H4**) is recovered? We need to have $R_{ww}(\tau) := E\{w(k+\tau)w^*(k)\} = 0$ for $|\tau| \neq 0$. By (3-9), $R_{ww}(\tau) = \mathbf{F}_0^{\mathcal{H}}R_{v_sv_s}(\tau)\mathbf{F}_0$. Define ($L > 0$ is some large integer)

$$\overline{R}_{v_sv_s} := \left[R_{v_sv_s}^T(-1) \ R_{v_sv_s}^T(-2) \ \cdots \ R_{v_sv_s}^T(-L)\right]^T \qquad (3-12)$$

where $R_{v_sv_s}(\tau) := E\{\mathbf{v}_s(k+\tau)\mathbf{v}_s^{\mathcal{H}}(k)\}$.

Lemma 3. $\overline{R}_{v_sv_s}$ is rank deficient for any $L \geq 1$ such that $\overline{R}_{v_sv_s}\mathbf{F}_0 = 0$. •

Proof: We have

$$R_{wv_s}(\tau) = E\{w(k+\tau)\mathbf{v}_s^{\mathcal{H}}(k)\} = 0 \quad \forall \tau \geq 1 \qquad (3-13)$$

because $\mathbf{v}_s(k)$ is obtained by causal filtering of $\mathbf{y}(k)$, hence of $w(k)$. Using (3-10) in (3-13) it then follows that there exists a $N \times 1$ $\mathbf{F}_0 \neq 0$ such that $\mathbf{F}_0^{\mathcal{H}}R_{v_sv_s}(\tau) = 0\ \forall \tau \geq 1$. Equivalently (since $R_{v_sv_s}(-\tau) = R_{v_sv_s}^{\mathcal{H}}(\tau)$)

$$R_{v_sv_s}(-\tau)\mathbf{F}_0 = 0 \quad \forall \tau \geq 1. \qquad (3-14)$$

The desired result is then immediate. □

Pick a $N \times 1$ column-vector $\mathbf{H}_0$ to equal the right-most right singular vector in a singular-value decomposition

(SVD) $\overline{R}_{v_s v_s} = U\Sigma V^{\mathcal{H}}$, i.e. the right singular vector corresponding to the smallest singular value. In other words, pick $\mathbf{H}_0$ to equal the last column of V. Then since ideally the smallest singular value of $\overline{R}_{v_s v_s}$ is zero, we have $\mathbf{H}_0^{\mathcal{H}} R_{v_s v_s}(\tau)\mathbf{H}_0 = 0$ for $\tau = 1, 2, \cdots, L$. Since the overall system with $w(k)$ as input and $\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k)$ as output is ARMA$(n_a, n_b + L_e)$, it follows that $\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k)$ is zero-mean white if $L \geq n_b + L_e$, hence, a scaled version of $w(k)$. Therefore, we have ($\alpha \neq 0$)

$$\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k) \;=:\; w'(k) = \alpha w(k) \qquad (3-15)$$

(because $\overline{R}_{v_s v_s}\mathbf{H}_0 = 0$). Thus, once $\mathbf{H}_0$ is found, one has the complete inverse filter to recover a scaled version of $w(k)$ via a zero-forcing filter.

Remark 2: $\mathbf{F}_0$ can also be estimated (up to a scale factor as unit norm $\mathbf{H}_0$) using the prediction error method of [9],[14] (even though [9] and [14] restrict their discussion to FIR models and real-valued data). Using (2-3) we obtain ($L_e \geq n_a + n_b - 1$)

$$[\ \mathbf{D}_1 \ \cdots \ \mathbf{D}_{L_e}\ ]\mathcal{R}_{ssL_e} = -[\ \mathbf{R}_{ss}(1) \ \cdots \ \mathbf{R}_{ss}(L_e)\ ] \qquad (3-16)$$

leading to the minimum norm solution

$$[\ \mathbf{D}_1 \ \cdots \ \mathbf{D}_{L_e}\ ] = -[\ \mathbf{R}_{ss}(1) \ \cdots \ \mathbf{R}_{ss}(L_e)\ ]\mathcal{R}_{ssL_e}^{\#}. \qquad (3-17)$$

Note that if $L_e > n_a + n_b - 1$, then $\mathbf{D}_i = 0$ for all $i > n_a + n_b - 1$ by Lemma 2. By (2-3)–(2-4) we have

$$\mathbf{R}_{II}(0) = \mathbf{F}_0\mathbf{F}_0^{\mathcal{H}} = \mathbf{R}_{ss}(0) + \sum_{i=1}^{L_e}\mathbf{D}_i\mathbf{R}_{ss}(-i). \qquad (3-18)$$

Clearly $\rho(\mathbf{R}_{ss}(0)) = 1$. Carry out an eigendecomposition of $\mathbf{R}_{II}(0)$. Pick $\mathbf{H}_0$ as the unit norm eigenvector corresponding to the largest eigenvalue (ideally the only nonzero eigenvalue) of $\mathbf{R}_{II}(0)$. □

Remark 3: It is worth noting that although $\mathbf{F}_0^{\mathcal{H}}\mathbf{v}_s(k) = w(k)$ (see (3-10)) and $\|\mathbf{F}_0\|^{-2}\mathbf{F}_0^{\mathcal{H}}I_s(k) = w(k)$ (see (2-4)), $\{I_s(k)\}$ is zero-mean white (linear innovations) whereas $\{\mathbf{v}_s(k)\}$ is in general colored. □

3.2. MMSE Equalizer with Delay d

We wish to design an MMSE linear equalizer of a specified length. Using the orthogonality principle [16], the MMSE equalizer of length L_e+1 to estimate $w(k-d)$ ($d \geq 0$) based upon $\mathbf{y}(n)$, $n = k, k-1, \cdots, k-L_e$, satisfies

$$[\ \overline{\mathbf{G}}_{d,0} \ \ \overline{\mathbf{G}}_{d,1} \ \cdots \ \overline{\mathbf{G}}_{d,L_e}\ ] =$$

$$[\ \mathbf{F}_d^{\mathcal{H}} \ \ \mathbf{F}_{d-1}^{\mathcal{H}} \ \cdots \ \mathbf{F}_0^{\mathcal{H}} \ \ 0 \ \cdots \ 0\ ]\mathcal{R}_{yyL_e}^{-1} \qquad (3-19)$$

where $\mathcal{R}_{yyL_e}$ has its ij–th block-element given by $\mathbf{R}_{yy}(j-i)$. Clearly one can obtain a consistent estimate of $\mathcal{R}_{yyL_e}$ from the given data. It remains to estimate $\mathbf{F}_l$'s to complete the design. Here the discussion of Sec. 3.1 becomes relevant. There we found a $\mathbf{H}_0$ to satisfy (3-15). From (3-9) and (3-15) we have

$$\mathbf{H}_0^{\mathcal{H}}\mathbf{v}_s(k) = \sum_{i=0}^{L_e}\mathbf{H}_0^{\mathcal{H}}\mathbf{G}_{ai}\mathbf{s}(n-i). \qquad (3-20)$$

Using (1-2), (3-15) and (3-20), we have

$$\mathbf{F}_\tau^{\mathcal{H}} = \alpha^{-1}\mathbf{H}_0^{\mathcal{H}}\sum_{i=0}^{L_e}\mathbf{G}_{ai}\mathbf{R}_{ss}^{\mathcal{H}}(\tau+i). \qquad (3-21)$$

Let $\mathcal{R}_{d,ssL_e}$ denote a $[N(L_e+1)]\times[N(L_e+1)]$ matrix with its ij-th block element as $E\{\mathbf{s}(k+d+j-i)\mathbf{s}^{\mathcal{H}}(k)\}$. Using (3-6) and (3-21) in (3-19) we obtain the desired solution

$$[\ \overline{\mathbf{G}}_{d,0} \ \ \overline{\mathbf{G}}_{d,1} \ \cdots \ \overline{\mathbf{G}}_{d,L_e}\ ]$$

$$= \alpha^{-1}\mathbf{H}_0^{\mathcal{H}}[\ I_{N\times N} \ \ 0 \ \cdots \ 0\ ]\mathcal{R}_{ssL_e}^{\#}\mathcal{R}_{d,ssL_e}^{\mathcal{H}}\mathcal{R}_{yyL_e}^{-1}. \qquad (3-22)$$

A scaled MMSE estimate of $w(t-d)$ is then given by

$$\widehat{w}(t-d) = \sum_{i=0}^{L_e}\alpha\overline{\mathbf{G}}_{d,i}\mathbf{y}(t-i). \qquad (3-23)$$

3.3. Summary of Algorithms

Given data $\mathbf{y}(k)$, $k = 1, 2, \cdots, T$. Pick the length $L_e + 1$ and delay d of the MMSE equalizer. Estimate all correlation functions by sample averaging.

3.3.1. ALGORITHM I :

Here $\mathbf{F}_0$ is estimated as the unit norm $\mathbf{H}_0$ that lies in the null space of $\overline{R}_{v_s v_s}$. Estimate noisefree correlations via (3-5). Use (3-22) and (3-23) for MMSE equalizer design.

3.3.2. ALGORITHM II :

Here $\mathbf{F}_0$ is estimated as in Remark 2. The rest is as in ALGORITHM I.

3.3.3. ALGORITHM III :

Here we will use (3-19) with $\mathbf{F}_i$ ($i = 0, 1, \cdots, d$) estimated using the basic approach of [9] and [14]. Although [9] and [14] derive all their results under the assumption of FIR channels with no common zeros, their results extend (with straightforward modifications) to models that satisfy (**H1**)-(**H5**) by virtue of Lemma 1.

4. BLIND EQUALIZATION: COMMON ZEROS

4.1. Minimum-Phase Zeros

Here the SIMO transfer function is

$$\mathcal{F}(z) = [\mathcal{B}_c(z)/\mathcal{A}(z)]\mathcal{B}(z) \qquad (4-1)$$

where $\mathcal{B}(z)$ satisfies (H2) and $\mathcal{B}_c(z)$ is a finite-degree scalar polynomial that collects all the common zeros of the subchannels. Assume that

(**H6**) Given model (4-1), $\mathcal{B}_c(z) \neq 0$ for $|z| \geq 1$.

Then while $\mathcal{A}^{-1}(z)\mathcal{B}(z)$ has a finite inverse, $\mathcal{B}_c^{-1}(z)$ is IIR though causal under (H6). Then (3-2) holds true approximately for "large" L_e, the approximation getting better with increasing L_e. Similarly Lemma 1 holds true approximately for "large" M and Lemma 2 also holds true approximately for $L_e \geq M$. It is then readily seen that the developments of Secs. 3.1, 3.2 and 3.3 are applicable.

4.2. Arbitrary Zeros

In this case (4-1) is true but $\mathcal{B}_c(z)$ does not necessarily satisfy (**H6**). We may rewrite (4-1) as

$$\mathcal{F}(z) = \overline{\mathcal{F}}(z)\mathcal{F}_{AP}(z) \qquad (4-2)$$

where $\mathcal{F}_{AP}(z)$ is an allpass (rational) function such that

$$\mathcal{B}_c(z)\mathcal{B}_c(z^{-1}) = \mathcal{F}_{AP}(z)\overline{\mathcal{B}}_{MP}(z) \qquad (4-3)$$

and $\overline{\mathcal{B}}_{MP}(z)$ is minimum-phase. Thus (within a scale factor) we have

$$\overline{\mathcal{F}}(z) = [\overline{\mathcal{B}}_{MP}(z)/\mathcal{A}(z)]\,\mathcal{B}(z). \qquad (4-4)$$

We may rewrite (1-2) as

$$\mathbf{y}(k) = \overline{\mathcal{F}}(z)w'(k) + \mathbf{n}(k) \text{ where } w'(k) := \mathcal{F}_{AP}(z)w(k). \quad (4-5)$$

Clearly $w'(k)$ satisfies (**H4**). Hence, (4-4)-(4-5) satisfy the requirements of Sec. 4.1. Therefore, one can "approximately" recover $w'(k)$ from the given data by applying the algorithms of Sec. **3.3**. In order to recover $w(k)$ form $w'(k)$, one needs to exploit the higher-order statistics of $\{w'(k)\}$; see [2],[3] and references therein.

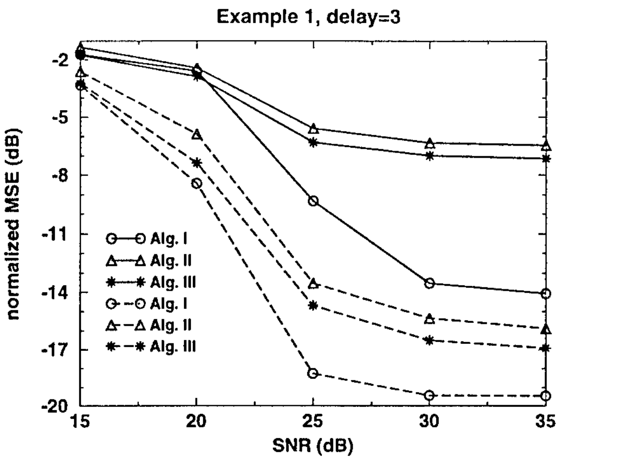


Fig. 1. Normalized MSE after MMSE equalization with $d = 3$. *Solid lines:* $T = 250$ *symbols, dashed lines:* $T = 1000$ *symbols.*

5. SIMULATION EXAMPLES

5.1. Example 1.

We have $N = 3$ in (1-2) with $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$ where

$$\mathcal{A}(z) = (1 - 0.5z^{-1})I_{3\times 3} \quad (5-1)$$

and $\mathcal{B}(z)$ is 3×1 MA(6) obtained from [10] as follows. Consider a raised cosine pulse $p_6(t, 0.1)$ with a roll-off factor 0.1, truncated to a length of $6T_s$ (T_s = symbol duration). As in [10], a two-ray multipath channel with (effective) impulse response $h(t) = p_6(t, 0.1) - 0.7p_6(t - T_s/3, 0.1)$ was sampled at intervals of $T_s/3$ (starting at $t = -3T_s$) to create the $\mathcal{B}(z)$ above. Transfer function $\mathcal{B}(z)$ satisfies (**H2**) [10], therefore, there exists a finite left inverse of length $L_e = 6$ (cf. Sec. 2.1). The scalar input $w(k)$ is 4-QAM. An MMSE equalizer of length $L_e = 8$ (9 taps per subchannel, totaling 27 taps —— overfitting) was designed with a delay d =3 (arbitrarily selected just for illustration). The Algorithms I–III were applied for record lengths $T = 250$ and 1000 symbols with varying SNR's. Fig. 1 shows the normalized MSE (MSE divided by $E\{|w(k)|^2\}$). It is seen that the proposed design approach can handle IIR channels with little difficulty. Algorithm I (newly proposed) performs the best.

5.2. Example 2.

Again we have $N = 3$ in (1-2) but with $\mathcal{F}(z) = \mathcal{B}_c(z)\mathcal{B}(z)$ where $\mathcal{B}(z)$ is as in Example 1 and $\mathcal{B}_c(z)$ is a scalar polynomial given by

$$\mathcal{B}_c(z) = 1 - 0.5z^{-1}. \quad (5-2)$$

Thus all three subchannels have a common zero at 0.5. The input $w(k)$ is 4-QAM as in Example 1. Note that in this example a finite left inverse does not exist. As in Example 1, an MMSE equalizer of length $L_e = 12$ was designed with a delay d =3. Fig. 2 shows the normalized MSE averaged over 100 Monte Carlo runs. It is seen that the proposed design approaches can handle subchannels with common minimum-phase zeros with little difficulty. As in Example 1, Algorithm I performs the best.

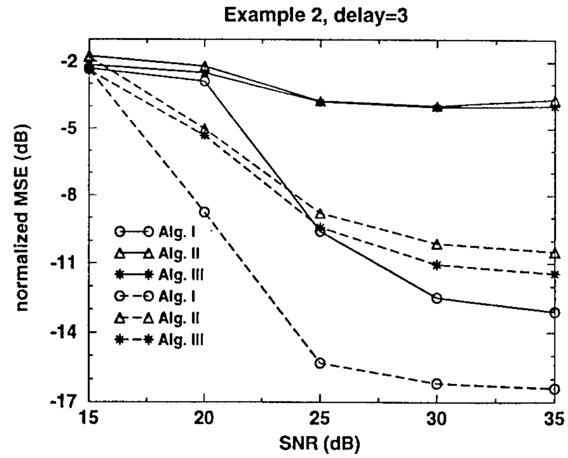


Fig. 2. Normalized MSE after MMSE equalization with $d = 3$. *Solid lines:* $T = 250$ *symbols, dashed lines:* $T = 1000$ *symbols.*

6. REFERENCES

[1] G.B. Giannakis and S.D. Halford, in *Proc. 1995 IEEE ICASSP*, pp. 1972-1975, Detroit, MI, May 9-12, 1995.

[2] J.K. Tugnait, *IEEE Trans. Signal Processing*, vol. SP-45, pp. 268-271, Jan. 1997.

[3] Special Issue, *IEEE Transactions on Signal Processing*, vol. SP-45, Jan. 1997.

[4] L. Tong, G. Xu and T. Kailath, *IEEE Trans. Information Theory*, vol. IT-40, pp. 340-349, March 1994.

[5] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 516-525, Feb. 1995.

[6] J.K. Tugnait, in *Proc. 30th Annual Asilomar Conf. Signals Systems Computers*, pp. 295-299, Pacific Grove, CA, Nov. 1996.

[7] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[8] D. Slock, in *Proc. 1994 IEEE ICASSP*, pp. IV:585-588, Adelaide, Australia, May 1994.

[9] K. Abed-Meraim et al. in *Proc. 1995 IEEE ICASSP*, pp. 1968-1971, Detroit, MI, May 9-12, 1995.

[10] Z. Ding, in *Proc. 1996 IEEE ICC*, pp. 852-856, Dallas, TX, June 23-26, 1996.

[11] K. Abed-Meraim et al. *IEEE Trans. Signal Processing*, vol. SP-45, pp. 42-55, Jan. 1997.

[12] D. Gesbert and P. Duhamel, in *Proc. Eighth IEEE Signal Processing Workshop on SSAP*, pp. 168-171, Corfu, Greece, June 24-26, 1996.

[13] D. Gesbert, P. Duhamel and S. Mayrargue, in *Proc. Eighth IEEE Signal Processing Workshop on SSAP*, pp. 172-175, Corfu, Greece, June 24-26, 1996.

[14] K. Abed-Meraim et al. *IEEE Trans. Signal Processing*, vol. SP-45, pp. 694-705, March 1997.

[15] D.G. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley, 1969.

[16] H.V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag: New York, 1988.

[9]

ADAPTIVE BLIND SEPARATION OF CONVOLUTIVE MIXTURES OF INDEPENDENT LINEAR SIGNALS

Jitendra K. Tugnait

Department of Electrical Engineering
Auburn University, Auburn, Alabama 36849, USA
E-mail: tugnait@eng.auburn.edu

ABSTRACT

This paper is concerned with the problem of blind separation of independent signals (sources) from their linear convolutive mixtures. The various signals are assumed to be linear non-Gaussian but not necessarily i.i.d. Recently an iterative, normalized higher-order cumulant maximization based approach was developed using the fourth-order normalized cumulants of the "beamformed" data. A byproduct of this approach is a decomposition of the given data at each sensor into its independent signal components. In this paper an adaptive implementation of the above approach is developed using a stochastic gradient approach. Some further enhancements including a Wiener filter implementation for signal separation and adaptive filter reinitialization are also provided. A computer simulation example is presented.

1. INTRODUCTION

Given noisy measurements $y_i(k)$, $(i = 1,2,\cdots,N)$, at time k at N sensors, let these measurements be a linear convolutive mixture of M source signals $x_j(k)$, $(j = 1,2,\cdots,M)$:

$$y_i(k) = \sum_{j=1}^{M} U_{ij}(z)x_j(k) + n_i(k), \quad i = 1,2,\cdots,N, \qquad (1-1)$$

$$\Longrightarrow \quad \mathbf{y}(k) = \mathcal{U}(z)\mathbf{x}(k) + \mathbf{n}(k), \qquad (1-2)$$

where $ij-$th element of $\mathcal{U}(z)$ is $U_{ij}(z)$, $\mathbf{y}(k) = [y_1(k)\vdots y_2(k)\vdots\cdots\vdots y_N(k)]^T$, similarly for $\mathbf{x}(k)$ and $\mathbf{n}(k)$, z^{-1} is both the backward-shift operator (i.e., $z^{-1}x(k) = x(k-1)$, etc.) as well as the complex variable in the $\mathcal{Z}-$transform, $x_j(k)$ is the j-th input at sampling time k, $y_i(k)$ is the i-th output, $n_i(k)$ is the additive Gaussian measurement noise, and $U_{ij}(z) := \sum_{l=0}^{\infty} u_{ij}(l)z^{-l}$ is the scalar transfer function with $x_j(k)$ as the input and $y_i(k)$ as the output. We allow all of the above variables to be complex-valued.

Suppose that we design a MIMO dynamic system $\mathcal{E}(z)$ with N inputs and M outputs such that the overall $M \times M$ system

$$\mathcal{T}(z) := \mathcal{E}(z)\mathcal{U}(z) \qquad (1-3)$$

decouples the source signals. Following the 2×2 case considered in [4], this implies that we must have ($T_{ij}(z)$ denotes the $ij-$th element of $\mathcal{T}(z)$)

$$T_{ij}(z) \begin{array}{ll} = 0 & \text{for} \quad i \neq i_j \\ \neq 0 & \text{for} \quad i = i_j \end{array} \qquad (1-4)$$

where $i = 1,2,\cdots,M$; $j = 1,2,\cdots,M$ and $i_j \in \{1,2,\cdots,M\}$ such that $i_j \neq i_l$ for $j \neq l$. That is, in every column and every row of $\mathcal{T}(z)$ there is exactly one non-zero entry. In a blind separation problem, the nonzero entries of $\mathcal{T}(z)$ are allowed to be a scalar linear system (shaping filter), unlike the equalization problems where they must be constant gains and/or pure delays.

The problem considered above arises in a wide variety of applications: array processing, speech enhancement ("cocktail party" problem), and noise cancellation, see [1]-[12] and references therein. The prior work done can be classified into two broad categories based upon the underlying propagation model: instantaneous mixtures and convolutive mixtures. The general model (1-2) represents a linear convolutive mixture. The work reported in [4], [7] and [11] (and references therein) deals with linear convolutive mixture (dynamic mixing) models. Past work on separation of convolutive mixtures may be categorized into several classes: time-domain approaches ([7], [8], [9], [10]), frequency-domain approaches ([4],[11]), adaptive (recursive) approaches ([7], [9], [10]) and non-recursive (batch) approaches ([4], [8], [11]). In this paper we present time-domain adaptive approaches. Quite a few of existing approaches are limited either to $M = N = 2$ ([4], [9]) or to $M = N$ ([7]). Although [11] treats a general case, their analysis is restricted to the case of two sources ($M = 2$). In this paper we consider a general case of $N \geq M$ with M arbitrary.

2. MODEL ASSUMPTIONS

We impose the following conditions on model (1-1)-(1-2):

(AS1) $N \geq M$ (at least as many outputs as inputs).

(AS2) The vector sequence $\{\mathbf{x}(k)\}$ is stationary, its various components are mutually independent, and $\mathcal{U}(z)$ is stable. Moreover, $\{\mathbf{x}(k)\}$ is linear, i.e.

$$\mathbf{x}(k) = \mathcal{V}(z)\mathbf{w}(k), \qquad (2-1)$$

where $\{\mathbf{w}(k)\}$ is a zero-mean, $M-$vector stationary non-Gaussian process, temporally i.i.d. and spatially independent, with nonzero fourth cumulants. Because of the mutual independence of the components of $\mathbf{x}(k)$, we take $\mathcal{V}(z)$ to be diagonal.

(AS3) Consider the composite system

$$\mathbf{y}(k) = \mathcal{F}(z)\mathbf{w}(k)+\mathbf{n}(k), \quad \text{with} \quad \mathcal{F}(z) := \mathcal{U}(z)\mathcal{V}(z). \qquad (2-2)$$

Assume that $\text{rank}\{\mathcal{F}(z)\} = M$ for any $|z| = 1$.

(AS4) Since the composite system is causal, we have

$$\mathcal{F}(z) = \sum_{l=0}^{\infty} \mathbf{F}_l z^{-l} \approx \sum_{l=0}^{L} \mathbf{F}_l z^{-l}. \qquad (2-3)$$

(AS5) The noise $\{\mathbf{n}(k)\}$ is a zero-mean, stationary Gaussian sequence independent of $\{\mathbf{w}(k)\}$.

Let $\mathcal{F}^{(i)}(z)$ denote the $i-$th column of $\mathcal{F}(z)$. In blind convolutive signal separation we are interested in decomposing the observations at the various sensors into its independent components. That is, our objective is to estimate $\mathcal{F}^{(i)}(z)w_i(k)$ for $i = 1,2,\cdots,M$ given $\{\mathbf{y}(k)\}$ without having a prior knowledge of $\mathcal{F}(z)$. Denote the $ij-$th element of $\mathcal{F}(z)$ as $F_{ij}(z)$.

This work was supported by NSF Grant MIP-9312559 and by ONR Grant N00014-97-1-0822.

3. A BATCH SOLUTION [8]

In this section we briefly discuss the batch (non-recursive) approach of [8]; its adaptive version is developed in Sec. 4. Let $\text{CUM}_4(w)$ denote the fourth-order cumulant of a complex-valued scalar zero-mean random variable w, defined as

$$\text{CUM}_4(w) = E\{|w|^4\} - 2[E\{|w|^2\}]^2 - |E\{w^2\}|^2. \quad (3-1)$$

Consider an $1 \times N$ row-vector polynomial equalizer (filter) $\mathcal{C}^T(z)$, with its j-th entry denoted by $\mathcal{C}_j(z)$, operating on the data vector $\mathbf{y}(k)$. Let the equalizer output be denoted by $e(k)$:

$$e(k) = \sum_{i=1}^{N} \mathcal{C}_i(z) y_i(k). \quad (3-2)$$

Following [6] consider maximization of the cost

$$J := \frac{|\text{CUM}_4(e(k))|}{[E\{|e(k)|^2\}]^2} \quad (3-3)$$

for designing a linear equalizer to recover one of the inputs. It is shown [6] that when (3-3) is maximized w.r.t. $\mathcal{C}(z)$, then (3-2) reduces to

$$e(k) = d w_{j_0}(k - k_0), \quad (3-4)$$

where d is some complex constant, k_0 is some integer, j_0 indexes some input out of the given M inputs.

An source-iterative solution is given by [8]:

Step 1. Maximize (3-3) w.r.t. $\mathcal{C}(z)$ to obtain (3-4).

Step 2. Cross-correlate $\{e(k)\}$ (of (3-4)) with the given data (2-2) and define a possibly scaled and shifted estimate of $f_{ij_0}(\tau)$ as

$$\widehat{f}_{ij_0}(\tau) := \frac{E\{y_i(k) e^*(k-\tau)\}}{E\{|e(k)|^2\}} \quad (3-5)$$

where $F_{ij}(z) = \sum_{l=-\infty}^{\infty} f_{ij}(l) z^{-l}$. Consider now the reconstructed contribution of $e(k)$ to the data $y_i(k)$ $(i = 1, 2, \cdots, N)$, denoted by $\widetilde{y}_{i,j_0}(k)$:

$$\widetilde{y}_{i,j_0}(k) := \sum_l \widehat{f}_{ij_0}(l) e(k-l). \quad (3-6)$$

Step 3. Remove the above contribution from the data to define the outputs of a MIMO system with N outputs and $M-1$ inputs. These are given by

$$y_i'(k) := y_i(k) - \widetilde{y}_{i,j_0}(k). \quad (3-7)$$

Step 4. If $M > 1$, set $M \leftarrow M-1$, $y_i(k) \leftarrow y_i'(k)$, and go back to **Step 1**, else quit.

It has been shown in [6],[8] that

$$\widetilde{y}_{i,j_0}(k) = \sum_l f_{ij_0}(l) w_{j_0}(k-l), \quad (3-8)$$

i.e., we have decomposed the observations at the various sensors into its independent components: $\widetilde{y}_{i,j_0}(k)$ in (3-8) represents the contribution of $\{w_{j_0}(k)\}$ to the i-th sensor achieving **blind signal separation**. It has been shown in [6] that under the conditions **(AS1)-(AS4)** and no noise, the proposed iterative approach is capable of blind identification of a MIMO transfer function $\mathcal{F}(z)$ up to a time-shift, a scaling and a permutation matrix provided that we allow doubly-infinite equalizers.

4. ADAPTIVE ALGORITHM

In this section we develop a stochastic gradient-based "recursification" of all of the batch optimization steps discussed in Sec. 3. We will use the superscript (m) to denote the various quantities pertaining to stage m of the batch algorithm of Sec. 3 (i.e. m-th execution of Steps 1-4 therein). Let the length of the equalizer $\mathcal{C}(z)$ be L_e and let

$$\mathcal{C}_i(z) = \sum_{l=0}^{L_e-1} c_i(l) z^{-l}. \quad (4-1)$$

Initialization: Define

$$Y_i(k) = [\; y_i(k) \quad \cdots \quad y_i(k-L_e+1) \;]^T, \quad (4-2)$$

$$\mathbf{Y}^{(1)}(k) = [\; Y_1^T(k) \quad \cdots \quad Y_N^T(k) \;]^T, \quad (4-3)$$

$$\mathbf{y}^{(1)}(k) = \mathbf{y}(k). \quad (4-4)$$

DO FOR $m = 1, 2, \cdots, M$:

$$\widetilde{\mathbf{C}}^{(m)}(k) = \mathbf{C}^{(m)}(k-1) + \mu_1 \nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k-1)) \quad (4-5)$$

$$\mathbf{C}^{(m)}(k) = \widetilde{\mathbf{C}}^{(m)}(k)/\|\widetilde{\mathbf{C}}^{(m)}(k)\| \quad (4-6)$$

where

$$\nabla_{\mathbf{C}^*} J_k^{(m)}(\mathbf{C}^{(m)}(k)) = \text{sgn}(\gamma_{4k}^{(m)}) \frac{2}{m_{2k}^{(m)3}} \times \Big\{ \Big[m_{2k}^{(m)} \left(e^{(m)2}(k) - \widetilde{m}_{2k}^{(m)} \right) e^{(m)*}(k) - \left(m_{4k}^{(m)} - |\widetilde{m}_{2k}^{(m)}|^2 \right) e^{(m)}(k) \Big] \mathbf{Y}^{(m)*}(k) \Big\}, \quad (4-7)$$

$$m_{2k}^{(m)} = (1-\mu_2) m_{2(k-1)}^{(m)} + \mu_2 |e^{(m)}(k)|^2, \quad (4-8)$$

$$\widetilde{m}_{2k}^{(m)} = (1-\mu_2) \widetilde{m}_{2(k-1)}^{(m)} + \mu_2 e^{(m)2}(k), \quad (4-9)$$

$$m_{4k}^{(m)} = (1-\mu_2) m_{4(k-1)}^{(m)} + \mu_2 |e^{(m)}(k)|^4, \quad (4-10)$$

$$\gamma_{4k}^{(m)} = m_{4k}^{(m)} - 2 m_{2k}^{(m)2} - |\widetilde{m}_{2k}^{(m)}|^2 \quad (4-11)$$

and

$$e^{(m)}(k) = \mathbf{C}^{(m)T}(k) \mathbf{Y}^{(m)}(k). \quad (4-12)$$

Set

$$\widehat{\widetilde{\mathbf{y}}}^{(m)}(k) = \sum_{n=-L_1}^{L_2} \widetilde{\mathbf{F}}_n^{(m)}(k) e^{(m)}(k-n) \quad (4-13)$$

where $\widehat{\widetilde{\mathbf{y}}}^{(m)}(k)$ represents (cf. (3-6)) the contribution of the extracted source at the m-th stage to the measurements at time k, and where $(n = -L_1, -L_1+1, \cdots, L_2)$

$$\widetilde{\mathbf{F}}_n^{(m)}(k) = \mathbf{R}_n^{(m)}(k)/m_{ee}^{(m)}(k), \quad (4-14)$$

$$m_{ee}^{(m)}(k) = (1-\mu_3) m_{ee}^{(m)}(k-1) + \mu_3 |e^{(m)}(k)|^2, \quad (4-15)$$

$$\mathbf{R}_n^{(m)}(k) = (1-\mu_3) \mathbf{R}_n^{(m)}(k-1) + \mu_3 \mathbf{y}^{(m)}(k) e^{(m)*}(k-n) \quad (4-16)$$

and

$$\mathbf{y}^{(m+1)}(k) = \mathbf{y}^{(m)}(k) - \widehat{\widetilde{\mathbf{y}}}^{(m)}(k). \quad (4-17)$$

Define

$$\widetilde{Y}_i^{(m)}(k) = [\; \widetilde{y}_i^{(m)}(k) \quad \cdots \quad \widetilde{y}_i^{(m)}(k-L_e+1) \;]^T \quad (4-18)$$

where $\widetilde{y}_i^{(m)}(k)$ denotes the $i-$th component of $\widetilde{\mathbf{y}}^{(m)}(k)$. Set

$$\mathbf{Y}^{(m+1)}(k) = [\ Y_1^{(m+1)T}(k) \quad \cdots \quad Y_N^{(m+1)T}(k)\ ]^T \qquad (4-19)$$

where

$$Y_i^{(m+1)}(k) = Y_i^{(m)}(k) - \widetilde{Y}_i^{(m)}(k). \qquad (4-20)$$

ENDDO

The sequence $\{\widehat{\widetilde{\mathbf{y}}}^{(m)}(k)\}$ in (4-13) represents the contribution of the extracted source at the $m-$th stage to the measurements at time k. Variable $e^{(m)}(k)$ in (4-12) corresponds to (3-2), $J_k^{(m)}$ in (4-5) corresponds to (3-3), and $\nabla_{\mathbf{C}^*} J_k^{(m)}$ is the instantaneous gradient, all at time k and stage m. In (4-5) μ_1 is the update step-size and in (4-8)-(4-10) and (4-15)-(4-16), μ_2 and μ_3, respectively, are the forgetting factors (> 0, < 1).

Running Cost. To monitor the convergence of the equalizers in various stages of the algorithm, it is useful to calculate a running cost with the sign. Let $\tilde{J}_k^{(m)}$ denote the running cost for the $m-$th stage at time k, given by

$$\tilde{J}_k^{(m)} = \frac{m_{4k}^{(m)} - |\widetilde{m}_{2k}^{(m)}|^2}{m_{2k}^{(m)2}} - 2 \qquad (4-21)$$

where $m_{2k}^{(m)}$, $\widetilde{m}_{2k}^{(m)}$ and $m_{4k}^{(m)}$ are computed as in (4-8)-(4-10) but with a smaller μ_2.

5. FURTHER MODIFICATIONS

5.1. MMSE Signal Separation

5.1.1. *Non-recursive Processing*

A by-product of the solutions of Secs. 3 and 4 is the estimates of the system/channel impulse response. These estimates can be used to design MMSE estimators of $\mathcal{F}^{(i)}(z)w_i(k)$ with a controlled delay d to obtain an "optimum" performance (ignoring any effects of additive noise on the channel estimates). Let $\mathbf{F}_l^{(i)}$ denote the $i-$th column of $\mathbf{F}_l$. We wish to design a linear MMSE filter (equalizer) of length L_e+1 to estimate $\widetilde{\mathbf{y}}^{(j)}(k-d)$ as $\widehat{\widetilde{\mathbf{y}}}^{(j)}(k-d)$ given $\mathbf{y}(l)$ for $l = k, k-1, \cdots, k-L_e+1$ where $d \geq 0$,

$$\widetilde{\mathbf{y}}^{(j)}(k) := \mathcal{F}^{(j)}(z)w_j(k) = \sum_{l=0}^{L} \mathbf{F}_l^{(j)} w_j(k-l), \quad (5-1)$$

$$\widehat{\widetilde{\mathbf{y}}}^{(j)}(k-d) := \sum_{i=0}^{L_e-1} \mathbf{G}_i \mathbf{y}(k-i). \qquad (5-2)$$

Using the orthogonality principle, the desired solution is given by

$$[\ \mathbf{G}_0 \quad \cdots \quad \mathbf{G}_{L_e-1}\ ] = \sigma_{wj}^2 [\ \mathbf{H}_d \quad \cdots \quad \mathbf{H}_{d-L_e}\ ] \mathcal{R}_{yy}^{-1} \qquad (5-3)$$

where $\mathcal{R}_{yy}$ denotes a $[NL_e] \times [NL_e]$ correlation matrix with $\mathbf{R}_{yy}(j-i)$ as its ij-th block element,

$$\mathbf{R}_{yy}(p) := E\{\mathbf{y}(t+p)\mathbf{y}^{\mathcal{H}}(t)\}, \quad \mathbf{H}_{d-p} := \sum_{k=0}^{L} \mathbf{F}_k^{(j)} \mathbf{F}_{k+d-p}^{(j)\mathcal{H}}. \qquad (5-4)$$

In practice, we replace all the unknowns by their estimates. Also we design the equalizer only up to a scale factor by omitting σ_{wj}^2 from (5-3).

Remark 1. *Selection of Delay d:* In designing (5-2) the delay d was pre-determined. One may choose to select d via exhaustive optimization as detailed below. The MMSE when (5-2) is used can be expressed as

$$\mathcal{J}(d) = \mathrm{tr} E\left\{\widetilde{\mathbf{y}}^{(j)}(k-d)\widetilde{\mathbf{y}}^{(j)\mathcal{H}}(k-d)\right\} - \mathcal{J}'(d) \quad (5-5)$$

where

$$\mathcal{J}'(d) := \sigma_{wj}^4 \,\mathrm{tr}\, \mathcal{H}\mathcal{R}_{yy}^{-1}\mathcal{H}^{\mathcal{H}}, \qquad (5-6)$$

$$\mathcal{H} := [\ \mathbf{H}_d \quad \mathbf{H}_{d-1} \quad \cdots \quad \mathbf{H}_{d-L_e}\ ]. \qquad (5-7)$$

Since the first term on the right-side of (5-5) is independent of d, minimizing $\mathcal{J}(d)$ w.r.t. d is equivalent to maximizing $\sigma_{wj}^{-4}\mathcal{J}'(d)$. $\square$

5.1.2. *Adaptive Implementation*

Note that $\mathcal{R}_{yy}^{-1}$ does not depend upon the stage m of the algorithm of Sec. 4. Its computation can easily be recursified by using the matrix inversion lemma: see Table 13.1 on p. 569 in [13]. Denote the data-based adaptive estimate of $\mathcal{R}_{yy}^{-1}$ at time k as $\mathcal{P}_{yy}(k)$. Let $\mathbf{H}_l^{(m)}(k)$ denote the estimate of $\mathbf{H}_l$ at stage m and time k of the multistage algorithm of Sec. 4. Note that $\widetilde{\mathbf{F}}_n^{(m)}(k)$ in (4-14) (see also (3-5)) denotes an estimate of $\mathbf{F}_n^{(i)}$ for some $i \in \{1, 2, \cdots, M\}$ (up to a scale factor and time shift). Therefore, from (5-2) and (5-5) we obtain the adaptive implementation at stage m; details are omitted.

5.2. Adaptive Filter Reinitialization

In the source-iterative (multistage) approaches of Secs. 3 and 4, any errors in cancelling the extracted sources from the preceding stages $l = 1, 2, \cdots, m-1$ affect the performance at stage m. The only stage that is immune to this phenomenon is stage $m = 1$. A possible solution to alleviate this error propagation from stage-to-stage is to use parallel stages where we still have M stages for M sources but they all operate directly on the given data record in parallel but with different initializations of the equalizers. The problem here is how to ensure that each stage converges to a distinct source. Here we propose to initialize the parallel stages using the results of the serial multistage implementation of Sec. 4 coupled with an MMSE solution similar to that of Sec. 5.1. For stage $m = 1$, there are no changes to the algorithm of Sec. 4. For stages $m \geq 2$, run the algorithm of Sec. 4 till the running cost (4-21) reaches a steady-state. Given the estimates of the subchannel impulse response at stage m, we can design an MMSE filter (in a fashion similar to Sec. 5.1.2) to estimate $w_j(k-d)$ given $\mathbf{y}(l)$ for $l = k, k-1, \cdots, k-L_e+1$. Let the extracted $w_j(k)$ at stage m be denoted by $w^{(m)}(k)$. Mimicking Sec. 5.1.2, a recursive MMSE solution at stage m and time k is given by

$$\widehat{w}^{(m)}(k-d) := \sum_{i=0}^{L_e-1} \overline{\mathbf{G}}_i^{(m)}(k)\mathbf{y}(k-i) \qquad (5-8)$$

where

$$[\ \overline{\mathbf{G}}_0^{(m)}(k) \quad \overline{\mathbf{G}}_1^{(m)}(k) \quad \cdots \quad \overline{\mathbf{G}}_{L_e-1}^{(m)}(k)\ ]$$
$$= [\ \widetilde{\mathbf{F}}_d^{(m)\mathcal{H}}(k) \quad \cdots \quad \widetilde{\mathbf{F}}_0^{(m)\mathcal{H}}(k) \quad 0 \quad \cdots \quad 0\ ]\, \mathcal{P}_{yy}(k). \qquad (5-9)$$

At stage m and time k, $\widehat{w}^{(m)}(k-d)$ is an MMSE estimate (with delay d) of $e^{(m)}(k)$ for the parallel implementation. Note that $\mathcal{C}(z) = \sum_{i=0}^{L_e-1} \overline{\mathbf{G}}_i^{(m)}(k)z^{-i}$ is the desired MMSE initializer.

6. SIMULATION EXAMPLE

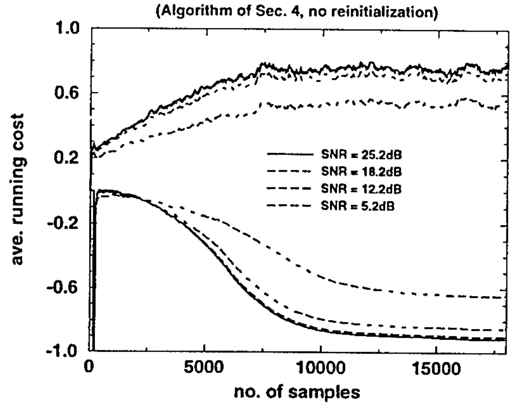
Take N=3 and M=2 in (2-2) with

$$\mathcal{F}^{(1)}(z) = \begin{bmatrix} 0.2 + 0.8z^{-1} + 0.4z^{-2} \\ 0.3z^{-1} - 0.6z^{-2} \\ 0. \end{bmatrix},$$

$$\mathcal{F}^{(2)}(z) = \begin{bmatrix} 0.5 - 0.3z^{-1} \\ -0.21z^{-1} - 0.5z^{-2} + 0.72z^{-3} + 0.36z^{-4} + 0.21z^{-6} \\ 0. \end{bmatrix}$$

Fig. 1.

(Algorithm of Sec. 4, no reinitialization)



The input $\{w_1(k)\}$ is an i.i.d. complex Gaussian-mixture with 4th normalized cumulant as 0.7433 . The input $\{w_2(k)\}$ is an i.i.d. 4-QAM sequence with 4th normalized cumulant as -1. The additive noise is white, complex Gaussian. The powers of $\{w_j(k)\}$ were scaled so as to have $E\{\|\mathcal{F}^{(1)}(z)w_1(k)\|^2\} = E\{\|\mathcal{F}^{(2)}(z)w_2(k)\|^2\}$. The performance measure was taken to be the signal-to-interference-and-noise ratio (SINR) per source signal, defined as

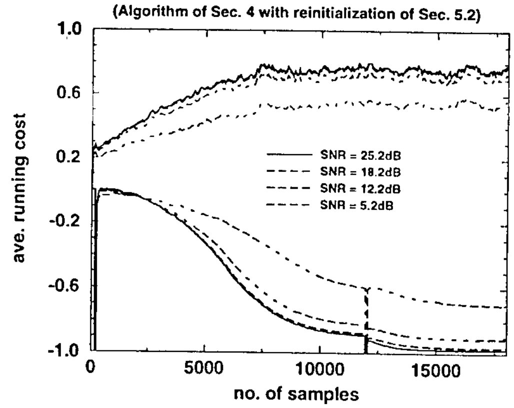
$$\mathrm{SINR}_j = \frac{E\{\|\widetilde{\mathbf{y}}^{(j)}(k)\|^2}{E\{\|\widetilde{\mathbf{y}}^{(j)}(k) - \widehat{\alpha}\widehat{\widetilde{\mathbf{y}}}^{(j)}(k)\|^2\}} \qquad (6-1)$$

where $\widehat{\alpha}$ is that value of the scalar α which minimizes $E\{\|\widetilde{\mathbf{y}}^{(j)}(k) - \alpha\widehat{\widetilde{\mathbf{y}}}^{(j)}(k)\|^2\}$. The length of the inverse filters was 11 samples per sensor (output) for the approach of Sec. 4. The initial guess for the tap gains was: set $c_i(5) = 1$ for $i = m$ for the $m-$th stage equalizer ($m = 1, 2$) with the remaining tap gains set to zero. The algorithm step sizes and forgetting factors for each stage m were chosen as: $\mu_1 = 0.0005$, $\mu_2 = 0.015$ and $\mu_3 = 0.0005$ when $\gamma_{4k}^{(m)} \leq 0$ (see (4-11)), and $\mu_1 = 0.00025$, $\mu_2 = 0.0075$ and $\mu_3 = 0.0005$ when $\gamma_{4k}^{(m)} > 0$. For the running cost (4-21) computation we selected "μ_2"=0.002 in (4-8)-(4-10). The parameters L_1 and L_2 in (4-13) were selected as $L_1 = 15$ and $L_2 = 6$. To design the MMSE equalizers/filters we took $L_e = 11$ and d was optimized following Remark 1 of Sec. 5.1.1 over the range $[-15, 6]$.

Fig. 1 shows the evolution of the average running cost $J_k^{(m)}$ (see (4-21)), averaged over 100 Monte Carlo runs (after 'assigning' each equalizer cost to its corresponding extracted source) without using any filter reinitialization. Fig. 2 shows $J_k^{(m)}$ when reinitialization (after 12000 samples) of Sec. 5.2 is used. It turns out that source 1 ($w_1(k)$) is extracted first, so that reinitialization only affects source 2 (4-QAM). Table I shows the average SINR (based on 100 runs) for the two sources at the end of the run (i.e. at k = 18000) without and with filter reinitialization, for various SNR's. The SINR's were computed using the solution (4-13) as well as the MMSE solution of Sec. 5.1.2. It is seen that blind signal separation benefits from both, MMSE signal separation as well as filter reinitialization.

Fig. 2.

(Algorithm of Sec. 4 with reinitialization of Sec. 5.2)



7. REFERENCES

[1] C. Jutten and J. Herault, *Signal Processing*, vol. 24, pp. 1-10, 1991.

[2] J.F. Cardoso and A. Souloumiac, *IEE Proc.-F, Radar and Signal Processing*, vol. 140, pp. 362-370, Dec. 1993.

[3] P. Comon, *Signal Processing*, vol. 36, No. 3, pp. 287-314, 1994.

[4] D. Yellin and E. Weinstein, *IEEE Trans. Signal Processing*, vol. SP-42, pp. 2158-2168, Aug. 1994.

[5] X.-R. Cao and R.-W. Liu, *IEEE Trans. Signal Proc.*, vol. SP-44, pp. 562-571, March 1996.

[6] J.K. Tugnait, *IEEE Trans. Signal Processing*, vol. SP-45, pp. 658-672, March 1997.

[7] P. Loubaton and P. Regalia, in *Proc. Intern. Conf. Commun.*, pp. 1160-1164, Geneva, Switzerland, June 1993.

[8] J.K. Tugnait, in *Proc. Eighth IEEE Signal Proc. Workshop on SSAP*, pp. 312-315, Corfu, Greece, June 24-26, 1996.

[9] U. Lindgren, T. Wigren and H. Broman, *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 3054-3058, Dec. 1995.

[10] N. Delfosse and P. Loubaton, in *Proc. 1996 ICASSP*, pp. 2940-2943, Atlanta, GA, May 7-10, 1996.

[11] C. Serviere, in *Proc. Eighth IEEE Signal Proc. Workshop on SSAP*, pp. 316-319, Corfu, Greece, June 24-26, 1996.

[12] J.F. Cardoso and B.H. Laheld, *IEEE Trans. Signal Proc.*, vol. SP-44, pp. 3017-3030, Dec. 1996.

[13] S. Haykin, *Adaptive Filter Theory*, 3rd Ed. Prentice-Hall: Upper Saddle River, NJ, 1996.

TABLE I. Average SINR after blind separation. Serial: Algorithm of Sec. 4; Parallel: Algorithm of Sec. 4 + reinitialization of Sec. 5.2.

SNR	SOURCE 1 (Gaussian mixture) serial		parallel	
	(4-13)	MMSE	(4-13)	MMSE
25.2 dB	8.653	10.667	8.653	10.667
18.2 dB	8.447	10.317	8.447	10.317
12.2 dB	7.807	9.253	7.807	9.253
5.2 dB	5.893	6.511	5.893	6.511

SNR	SOURCE 2 (4-QAM) serial		parallel	
	(4-13)	MMSE	(4-13)	MMSE
25.2 dB	11.621	12.647	16.123	15.271
18.2 dB	11.198	12.134	15.078	14.351
12.2 dB	9.876	10.591	12.445	12.070
5.2 dB	6.505	6.862	7.746	7.626

[10]

ON MULTI-STEP LINEAR PREDICTORS FOR M.I.M.O. F.I.R./I.I.R. CHANNELS AND RELATED BLIND EQUALIZATION

Jitendra K. Tugnait

Department of Electrical Engineering
Auburn University, Auburn, Alabama 36849, USA

ABSTRACT

In several communications (and related) applications the underlying equivalent discrete-time mathematical model is that of a multiple-input multiple-output (MIMO) linear system where the number of inputs equals the number of users (sources) and the number of outputs is related to the number of sensors and the sampling rate. The vector input sequence represents the information sequences of the various users. Existence of finite-length multi-step (including one-step) linear predictors plays a key role in blind identification and equalization of multiple-input multiple-output (MIMO) systems. In this paper we first derive an upper bound on the length of a linear predictor for MIMO systems with irreducible transfer functions. Then multi-step linear predictors for IIR/FIR MIMO channels are considered. An upper bound on the length of the one-step predictor is known for the case when the underlying MIMO transfer function is irreducible and column-reduced. When the MIMO transfer function is irreducible but not necessarily column-reduced, it is known that a finite-length linear predictor exists; however, its length has not been previously specified in the literature. In past multi-step linear predictors have been considered in the literature only for single-input multiple-output models.

1. INTRODUCTION

Consider a discrete-time IIR MIMO system with N outputs and M inputs:

$$\mathbf{y}(k) = \mathcal{F}(z)\mathbf{w}(k)+\mathbf{n}(k) = \mathbf{s}(k)+\mathbf{n}(k) \qquad (1)$$

where $\mathbf{y}(k)=[y_1(k)\vdots y_2(k)\vdots\cdots\vdots y_N(k)]^T$, similarly for $\mathbf{w}(k)$, $\mathbf{s}(k)$ and $\mathbf{n}(k)$, z is the $\mathcal{Z}-$transform variable as well as the backward-shift operator (i.e., $z^{-1}\mathbf{w}(k)=\mathbf{w}(k-1)$, etc.), $\mathbf{s}(k)$ is the noise-free output, $\mathbf{n}(k)$ is the additive measurement noise and the $N\times M$ matrix $\mathcal{F}(z)$ is given by

$$\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$$

where

$$\mathcal{A}(z)=I+\sum_{i=1}^{n_a}\mathbf{A}_i z^{-i} \text{ and } \mathcal{B}(z) = \sum_{i=0}^{n_b}\mathbf{B}_i z^{-i}. \qquad (2)$$

We allow all of the above variables to be complex-valued. The following assumptions are made regarding (1)-(2):

(H1) $N>M$.

(H2) Rank$\{\mathcal{B}(z)\}=M$ $\forall z$ including $z=\infty$ but excluding $z=0$, i.e., $\mathcal{B}(z)$ is irreducible [5, Sec. 6.3].

(H3) Unobserved input sequence $\{\mathbf{w}(k)\}$ is zero-mean, white. Take $E\{\mathbf{w}(k)\mathbf{w}^H(k)\}=I_M$ by absorbing any non-identity correlation of $w(k)$ into $\mathcal{B}(z)$ where I_M is the $M\times M$ identity matrix and the superscript H is the Hermitian operator (complex conjugate transpose).

(H4) $\{\mathbf{n}(k)\}$ is zero-mean with $E\{\mathbf{n}(k+\tau)\mathbf{n}^H(k)\}=\sigma_n^2 I_N\delta(\tau)$.

(H5) $\mathcal{A}(z)\neq 0$ for $|z|\geq 1$.

Notations and Definitions: Let $\mathcal{B}^{(l)}(z)$ denote the $l-$th column of $\mathcal{B}(z)$ such that $\mathcal{B}^{(l)}(z)=\sum_{i=0}^{L_l}\mathbf{B}_i^{(l)}z^{-i}$ where $L_l=\deg\left(\mathcal{B}^{(l)}(z)\right)=$ lowest degree of the polynomial $\mathcal{B}^{(l)}(z)$. By (2), $L_l\leq n_b$ $\forall l$. The polynomial matrix $\mathcal{B}(z)$ is said to be column-reduced if rank $\left\{[\mathbf{B}_{L_1}^{(1)}\vdots\cdots\vdots\mathbf{B}_{L_M}^{(M)}]\right\}=M$ [5]. Consider the Hilbert space $\mathcal{H}$ of square integrable complex random variables on a common probability space endowed with the inner product (for scaler complex random variables x_1 and x_2) $<x_1,x_2>=E\{x_1x_2^*\}$ where the superscript $*$ denotes complex conjugation (see [4]). Let $Sp\{x_i\in I\}$ denote the subspace of $\mathcal{H}$ generated by the random variables/vectors in the set $\{x_i\in I\}$. Given an $N-$variate $\mathbf{s}(k)$ with $i-$th component $s_i(k)$, define the subspace

$$H_{k-m;L_1,L_2,\cdots,L_N}(\mathbf{s}):=$$
$$Sp\{\ s_i(k-l_i),\ \ m\leq l_i\leq L_i;\ \ i=1,2,\cdots,N\}.$$

We will use $H_{k-m}(\mathbf{s})$ to denote $H_{k-m;\infty,\cdots,\infty}(\mathbf{s})$. Let $(\mathbf{s}(k)|H_{k-1}(\mathbf{s}))$ denote the orthogonal projection of $\mathbf{s}(k)$ onto the subspace $H_{k-1}(\mathbf{s})$ [4]. □

Models such as (1)-(2) with $\mathcal{F}(z)=\mathcal{B}(z)$ arise in several useful digital communications and other applications [1]-[3], [6]-[8] where one of the objectives is to estimate the multichannel impulse response $\{\mathbf{B}_i\}$ and/or to recover the inputs $\mathbf{w}(k)$ given the noisy measurements but not given the knowledge of the system transfer function. One of the popular approaches is that using linear prediction [1]-[3] where existence of finite-length one-step linear predictors plays a key role. In the MIMO case it is known that under (H1)-(H3), finite-length one-step linear predictors exist for the process $\mathbf{s}(k)$ [6],[7]. The length (or an upper-bound on it) has not been specified in [6],[7]. Under an additional condition that $\mathcal{B}(z)$ is column-reduced, it is stated in [1] that there exists a linear predictor (for $\mathbf{s}(k)$) of length no longer than $\sum_{i=1}^M L_i$.

A one-step linear prediction-based approach was first proposed in [12] and later expanded upon in [2]. Unlike the subspace-based methods of [13], [14] and others (see also

This work was supported by the Office of Naval Research under Grant N00014-97-1-0822.

[3] and references therein), the linear prediction (LP) based approach of [12] and [2] turns out to be rather insensitive to the order of the underlying FIR channel (so long as one overfits). More recently, it has been pointed out in [15] and [16] that the LP-based approach can be further significantly improved by utilizing some additional information not exploited by LP. Although [15] and [16] derive their algorithms in a quite a different manner, their final algorithms are essentially the same. In this paper we will follow the approach of [16] which is based upon multi-step linear prediction. Unlike [16] we allow multiple inputs and IIR channels. Unlike [15] we allow MIMO transfer functions that are not column-reduced and we also allow IIR channels.

2. FINITE-LENGTH ONE-STEP LINEAR PREDICTORS FOR $\mathcal{F}(Z) = \mathcal{B}(Z)$

By [5, Sec. 6.3] there exists an $M \times M$ unimodular matrix $\mathcal{W}(z)$ such that

$$\mathcal{B}(z) = \overline{\mathcal{B}}(z)\mathcal{W}(z) \tag{3}$$

where $\overline{\mathcal{B}}(z)$ is column-reduced and $\mathcal{W}(z)$ is unimodular (i.e. $\det(\mathcal{W}(z)) =$ constant). Let $\overline{\mathcal{B}}^{(l)}(z) = l$–th column of $\overline{\mathcal{B}}(z)$, $\overline{L}_l = \deg\left(\overline{\mathcal{B}}^{(l)}(z)\right)$ and $\overline{\mathcal{B}}^{(l)}(z) = \sum_{i=0}^{\overline{L}_l} \overline{\mathbf{B}}_i^{(l)} z^{-i}$. Let $L^{(W)} = \deg(\mathcal{W}(z))$. Then

$$\mathcal{W}^{-1}(z) = \sum_{i=0}^{p} \overline{W}_i z^{-i} \quad \text{where} \quad p \le (M-1)L^{(W)}. \tag{4}$$

Define the $[NK] \times [K + \overline{L}_i]$ generalized Sylvester matrix (a Toeplitz matrix)

$$\mathcal{T}_K(\overline{\mathcal{B}}^{(i)}) = \begin{bmatrix} \overline{\mathbf{B}}_0^{(i)} & \overline{\mathbf{B}}_1^{(i)} & \cdots & \overline{\mathbf{B}}_{\overline{L}_i}^{(i)} & 0 & \cdots & 0 \\ 0 & \overline{\mathbf{B}}_0^{(i)} & \cdots & \overline{\mathbf{B}}_{\overline{L}_i-1}^{(i)} & \overline{\mathbf{B}}_{\overline{L}_i}^{(i)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{\mathbf{B}}_0^{(i)} & \overline{\mathbf{B}}_1^{(i)} & \cdots & \overline{\mathbf{B}}_{\overline{L}_i}^{(i)} \end{bmatrix} \tag{5}$$

Further define the $[NK] \times [MK + \sum_{i=1}^{M} \overline{L}_i]$ matrix

$$\mathcal{T}_K(\overline{\mathcal{B}}) := \left[\begin{array}{cccc} \mathcal{T}_K(\overline{\mathcal{B}}^{(1)}) & \mathcal{T}_K(\overline{\mathcal{B}}^{(2)}) & \cdots & \mathcal{T}_K(\overline{\mathcal{B}}^{(M)}) \end{array}\right]. \tag{6}$$

Set $\mathbf{x}(k) := \mathcal{W}(z)\mathbf{w}(k)$ so that $\mathbf{s}(k) = \overline{\mathcal{B}}(z)\mathbf{x}(k)$. Then we have

$$S_K(k) = \mathcal{T}_K(\overline{\mathcal{B}})X_K(k) \tag{7}$$

where $S_K(k) := [\mathbf{s}^T(k) \vdots \cdots \vdots \mathbf{s}^T(k-K+1)]^T$ and $X_K(k) := [x_1(k) \vdots \cdots \vdots x_1(k - \overline{L}_1 - K + 1) \vdots x_2(k) \vdots \cdots \vdots x_2(k - \overline{L}_2 - K + 1) \vdots \cdots \vdots x_M(k) \vdots \cdots \vdots x_M(k - \overline{L}_M - K + 1)]^T$ ($x_i(k)$ is the i–th component of $\mathbf{x}(k)$). It is known [9] (see also [5],[8]) that $\operatorname{rank}\{\mathcal{T}_K(\overline{\mathcal{B}})\} = MK + \sum_{i=1}^{M} \overline{L}_i$ if $K \ge \sum_{i=1}^{M} \overline{L}_i$. Therefore, a left inverse to $\mathcal{T}_K(\overline{\mathcal{B}})$ exists. Hence, if $K \ge \sum_{i=1}^{M} \overline{L}_i$, we have

$$H_{k;K-1,K-1,\cdots,K-1}(\mathbf{s}) = H_{k;K+\overline{L}_1-1,K+\overline{L}_2-1,\cdots,K+\overline{L}_M-1}(\mathbf{x}) \tag{8}$$

Since $\mathbf{x}(k) = \mathcal{W}(z)\mathbf{w}(k)$, using (4) it follows that $\mathbf{w}(k) = \sum_{i=0}^{p} \overline{W}_i \mathbf{x}(k-i)$. Therefore, for some $\mathbf{C}_m$'s, we have

$$\widehat{\mathbf{s}}(k|k-1) := \sum_{i=1}^{n_b} \mathbf{B}_i \mathbf{w}(k-i) = \sum_{i=1}^{n_b} \mathbf{B}_i \left[\sum_{l=0}^{p} \overline{W}_l \mathbf{x}(k-i-l)\right]$$

$$= \sum_{m=1}^{p+n_b} \mathbf{C}_m \mathbf{x}(k-m). \tag{9}$$

Define $K_0 := \max\left\{\sum_{i=1}^{M} \overline{L}_i,\, n_b + p\right\}$. If $K \ge K_0$, then $\widehat{\mathbf{s}}(k|k-1) \in H_{k-1;p+n_b,\cdots,p+n_b}(\mathbf{x}) \subset H_{k-1;K+\overline{L}_1,\cdots,K+\overline{L}_M}(\mathbf{x}) = H_{k-1;K,\cdots,K}(\mathbf{s})$. Therefore, there exist $\overline{\mathbf{A}}_i$'s such that

$$\widehat{\mathbf{s}}(k|k-1) = -\sum_{i=1}^{K_0} \overline{\mathbf{A}}_i \mathbf{s}(k-i). \tag{10}$$

Using (1), (2) and (9), we have

$$\mathbf{s}(k) = \widehat{\mathbf{s}}(k|k-1) + \mathbf{e}(k|k-1) \quad \text{where} \quad \mathbf{e}(k|k-1) := \mathbf{B}_0\mathbf{w}(k). \tag{11}$$

By **(H3)** it follows that $E\{\mathbf{e}(k|k-1)\mathbf{s}^H(k-m)\} = 0\ \forall m \ge 1$. Hence, by the orthogonal projection theorem (OPT) [4], we have $\widehat{\mathbf{s}}(k|k-l) = (\mathbf{s}(k)|H_{k-l}(\mathbf{s}))$. But by (10) and OPT, we also have $\widehat{\mathbf{s}}(k|k-l) = (\mathbf{s}(k)|H_{k-1;K_0,\cdots,K_0}(\mathbf{s}))$. Thus, $\mathbf{e}(k|k-1)$ is the linear innovations process of $\{\mathbf{s}(k)\}$ [4]. It remains to 'simplify' K_0. By (3) and [5, Thm. 6.3-13] (the predictable-degree property of column-reduced matrices), it follows that $L^{(W)} \le n_b$ and $\overline{L}_l \le n_b$ $(1 \le l \le M)$. Therefore, $p \le (M-1)n_b$ and $\sum_{i=1}^{M} \overline{L}_i \le Mn_b$. Hence, $K_0 \le Mn_b$. The above discussion is summarized below.

$\boxed{\textbf{Theorem 1.}}$ Under **(H1)**-**(H3)**, there exists an integer $K \le Mn_b$ and a polynomial matrix $\overline{\mathcal{A}}(z) = I_N + \sum_{i=1}^{K} \overline{\mathbf{A}}_i z^{-i}$ of degree K such that $\overline{\mathcal{A}}(z)\mathbf{s}(k) = \mathbf{e}(k|k-1) = \mathbf{B}_0\mathbf{w}(k)$. The linear innovations process of $\{\mathbf{s}(k)\}$ is $\{\mathbf{e}(k|k-1)\}$. •

It follows from (1) and Theorem 1 that

$$\overline{\mathcal{A}}(z)\mathbf{s}(k) = \overline{\mathcal{A}}(z)\mathcal{B}(z)\mathbf{w}(k) = \mathbf{B}_0\mathbf{w}(k). \tag{12}$$

Since $\mathbf{w}(k)$ is full-rank and white, it follows that

$$\overline{\mathcal{A}}(z)\mathcal{B}(z) = \mathbf{B}_0 \quad \Rightarrow \quad \left(\mathbf{B}_0^H\mathbf{B}_0\right)^{-1}\mathbf{B}_0^H\overline{\mathcal{A}}(z)\mathcal{B}(z) = I_M. \tag{13}$$

Clearly the $M \times N$ polynomial matrix $\mathcal{G}(z) := \left(\mathbf{B}_0^H\mathbf{B}_0\right)^{-1}\mathbf{B}_0^H\overline{\mathcal{A}}(z)$ is of degree $K \le Mn_b$ and it is a left inverse to $\mathcal{B}(z)$.

$\boxed{\textbf{Theorem 2.}}$ Under **(H1)**-**(H3)**, there exists an integer $K \le Mn_b$ and a polynomial matrix $\mathcal{G}(z) = \sum_{i=0}^{K} \mathbf{G}_i z^{-i}$ of degree K such that $\mathcal{G}(z)\mathcal{B}(z) = I_M$. •

In [8] we derived the upper bound on $\deg(\mathcal{G}(z))$ as $(2M-1)n_b - 1$. Clearly Theorem 2 offers a better bound for $M \ge 2$.

3. FINITE-LENGTH MULTI-STEP LINEAR PREDICTORS

We now treat the general case $\mathcal{F}(z) = \mathcal{A}^{-1}(z)\mathcal{B}(z)$. We have

$$\mathcal{F}(z) = \sum_{i=0}^{\infty} \mathbf{F}_i z^{-i} \tag{14}$$

and

$$\mathbf{s}(k) = -\sum_{i=1}^{n_a} \mathbf{A}_i \mathbf{s}(k-i) + \sum_{i=0}^{n_b} \mathbf{B}_i \mathbf{w}(k-i). \tag{15}$$

It then follows from (15) that

$$\mathbf{s}(k) = -\sum_{i=2}^{n_a}\mathbf{A}_i\mathbf{s}(k-i) - \mathbf{A}_1\left[-\sum_{i=1}^{n_a}\mathbf{A}_i\mathbf{s}(k-1-i) + \sum_{i=0}^{n_b}\mathbf{B}_i\mathbf{w}(k-1-i)\right] + \sum_{i=0}^{n_b}\mathbf{B}_i\mathbf{w}(k-i)$$
$$= -\sum_{i=2}^{n_a+1}\mathbf{A}_{2i}\mathbf{s}(k-i) + \sum_{i=0}^{n_b+1}\mathbf{B}_{2i}\mathbf{w}(k-i). \quad (16)$$

for some appropriate choices of the parameters (matrices) $\mathbf{A}_{2i}$'s and $\mathbf{B}_{2i}$'s. Now substitute for $\mathbf{s}(k-2)$ using (15) in (16), and continuing this way, we have, in general, for appropriate choices of $\mathbf{A}_{li}$'s and $\mathbf{B}_{li}$'s ($l \geq 1$)

$$\mathbf{s}(k) = -\sum_{i=l}^{n_a+l-1}\mathbf{A}_{li}\mathbf{s}(k-i) + \sum_{i=0}^{n_b+l-1}\mathbf{B}_{li}\mathbf{w}(k-i). \quad (17)$$

Both (15) and (17) represent the same signal/system and therefore, they must have the same impulse response. By (14), (15) and (17), it then follows that

$$\mathbf{B}_{li} = \mathbf{F}_i \quad \text{for} \quad 0 \leq i \leq l-1. \quad (18)$$

Let us rewrite (17) as

$$\mathbf{s}(k) = \mathbf{e}(k|k-l) + \widehat{\mathbf{s}}(k|k-l) \quad (19)$$

where

$$\mathbf{e}(k|k-l) := \sum_{i=0}^{l-1}\mathbf{B}_{li}\mathbf{w}(k-i) = \sum_{i=0}^{l-1}\mathbf{F}_i\mathbf{w}(k-i) \quad (20)$$

and

$$\widehat{\mathbf{s}}(k|k-l) := -\sum_{i=l}^{n_a+l-1}\mathbf{A}_{li}\mathbf{s}(k-i) + \sum_{i=l}^{n_b+l-1}\mathbf{B}_{li}\mathbf{w}(k-i). \quad (21)$$

Theorem 3. Under **(H1)**-**(H3)**, **(H5)**, and for $l = 1, 2, \cdots$, $\{\mathbf{s}(k)\}$ can be decomposed as in (19) such that

$$E\{\mathbf{e}(k|k-l)\mathbf{s}^H(k-m)\} = 0 \quad \forall m \geq l, \quad (22)$$

$$\widehat{\mathbf{s}}(k|k-l) = (\mathbf{s}(k)|H_{k-l}(\mathbf{s})), \quad (23)$$

$$\widehat{\mathbf{s}}(k|k-l) \in H_{k-l;n_a+Mn_b+l-1,\cdots,n_a+Mn_b+l-1}(\mathbf{s}) \quad (24)$$

and

$$\widehat{\mathbf{s}}(k|k-l) = (\mathbf{s}(k)|H_{k-l;n_a+Mn_b+l-1,\cdots,n_a+Mn_b+l-1}(\mathbf{s})) \quad \bullet \quad (25)$$

Proof: By (1), (2), (14) and **(H5)**, we have

$$\mathbf{s}(k) = \sum_{i=0}^{\infty}\mathbf{F}_i\mathbf{w}(k-i). \quad (26)$$

By Theorem 2, it follows that

$$\sum_{i=0}^{Mn_b}\mathbf{G}_i\mathbf{s}(k-i) = \mathbf{w}(k). \quad (27)$$

Substituting for $\mathbf{w}(k)$ from (27) in (21), it follows that

$$\widehat{\mathbf{s}}(k|k-l) \in H_{k-l}(\mathbf{s}). \quad (28)$$

By (26) and **(H3)**, we have

$$E\{\mathbf{w}(k)\mathbf{s}^H(k-m)\} = 0 \quad \forall m > 0. \quad (29)$$

Therefore, using (20) and (29), it follows that (22) is true. By (19), (22), (28) and the orthogonal projection theorem [4], it follows that (23) is true (as the "error" $\mathbf{e}(k|k-l)$ is orthogonal to the data $\mathbf{s}(k-m)$ $(m \geq l)$, hence to the subspace $H_{k-l}(\mathbf{s})$).

It remains to establish (24) and (25). Define

$$\tilde{\mathbf{s}}(k) := \mathcal{A}(z)\mathbf{s}(k) = \mathcal{B}(z)\mathbf{w}(k)$$
$$= \overline{\mathcal{B}}(z)[\mathcal{W}(z)\mathbf{w}(k)] = \overline{\mathcal{B}}(z)\mathbf{w}(k) \quad (30)$$

where we have used (3). Using (30) and rewriting (8) in the notation of Sec. 3, if $K \geq \sum_{i=1}^{M}\overline{L}_i$, we have

$$H_{k;K-1,\cdots,K-1}(\tilde{\mathbf{s}}) = H_{k;K+\overline{L}_1-1,\cdots,K+\overline{L}_M-1}(\mathbf{x}). \quad (31)$$

It also follows from (30) that

$$H_{k;L,\cdots,L}(\tilde{\mathbf{s}}) \subset H_{k;n_a+L,\cdots,n_a+L}(\mathbf{s}) \quad \forall L \geq 0. \quad (32)$$

Therefore, for $K \geq \sum_{i=1}^{M}\overline{L}_i$,

$$H_{k;K+\overline{L}_1-1,\cdots,K+\overline{L}_M-1}(\mathbf{x}) \subset H_{k;n_a+K-1,\cdots,n_a+K-1}(\mathbf{s}), \quad (33)$$

and in general, for any $l \geq 0$,

$$H_{k-l;K+\overline{L}_1+l-1,\cdots,K+\overline{L}_M+l-1}(\mathbf{x})$$
$$\subset H_{k-l;n_a+K+l-1,\cdots,n_a+K+l-1}(\mathbf{s}). \quad (34)$$

As in (9) we have

$$\sum_{i=l}^{n_b+l-1}\mathbf{B}_{li}\mathbf{w}(k-i) = \sum_{i=l}^{p+n_b+l-1}\mathbf{C}_i\mathbf{x}(k-i) \quad (35)$$

for some $\mathbf{C}_m$'s. Therefore, it follows that

$$\sum_{i=l}^{n_b+l-1}\mathbf{B}_{li}\mathbf{w}(k-i) \in H_{k-l;p+n_b+l-1,\cdots,p+n_b+l-1}(\mathbf{x}) \quad (36)$$

$$\subset H_{k-l;K+\overline{L}_1+l-1,\cdots,K+\overline{L}_M+l-1}(\mathbf{x}) \quad \text{for } K \geq K_0 \quad (37)$$

$$\subset H_{k-l;n_a+K+l-1,\cdots,n_a+K+l-1}(\mathbf{s}) \quad (38)$$

where, as in Sec. 2, $K_0 := \max\left\{\sum_{i=1}^{M}\overline{L}_i,\ n_b+p\right\}$. It therefore follows from (21) and (38) that

$$\widehat{\mathbf{s}}(k|k-l) \in H_{k-l;n_a+K+l-1,\cdots,n_a+K+l-1}(\mathbf{s}) \quad \forall K \geq K_0. \quad (39)$$

As in Sec. 2, $K_0 \leq Mn_b$. If we pick $K = Mn_b$ in (39), we obtain (24). Finally, (25) follows from (19), (22), (24) and the orthogonal projection theorem [4]. This completes the proof of Theorem 3. □

It follows from Theorem 3 that

$$\widehat{\mathbf{s}}(k|k-l) = -\sum_{i=l}^{L_l}\overline{\mathbf{A}}_{li}\mathbf{s}(k-i) \quad \text{where} \quad L_l \geq n_a + Mn_b + l - 1, \quad (40)$$

for some $N \times N$ matrices $\overline{\mathbf{A}}_{li}$s. By (19) and (22) (recall also the orthogonal projection theorem), we have

$$\widehat{\mathbf{s}}(k|k-l) = \arg\left\{\min_{\mathbf{x}(k)\in H_{k-l}(\mathbf{s})} E\{\|\mathbf{s}(k)-\mathbf{x}(k)\|^2\right\}. \tag{41}$$

Therefore, $\widehat{\mathbf{s}}(k|k-l)$ is the l−step (ahead) linear predictor of $\mathbf{s}(k)$ given $\{\mathbf{s}(m),\ m \leq k-l\}$. By (25) it is also the l−step (ahead) linear predictor of $\mathbf{s}(k)$ given $\{\mathbf{s}(m),\ k-L_l \leq m \leq k-l\}$.

It follows from (19) and (40) that

$$\mathbf{e}(k|k-l) := \mathbf{s}(k) + \sum_{i=l}^{L_l} \overline{\mathbf{A}}_{li}\mathbf{s}(k-i) = \sum_{i=0}^{l-1} \mathbf{F}_i\mathbf{w}(k-i). \tag{42}$$

It follows from (42) that for $l \geq 2$,

$$\mathbf{e}_{d,l}(k) := \mathbf{e}(k|k-l) - \mathbf{e}(k|k-l+1) = \mathbf{F}_{l-1}\mathbf{w}(k-l+1). \tag{43}$$

Define a $[N \times D]$−vector $(D \geq 1)$

$$\mathbf{S}(k) := \begin{bmatrix} \mathbf{e}(k|k-1) \\ \mathbf{e}_{d,2}(k+1) \\ \vdots \\ \mathbf{e}_{d,D}(k+D-1) \end{bmatrix} = \begin{bmatrix} \mathbf{F}_0 \\ \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_{D-1} \end{bmatrix} \mathbf{w}(t). \tag{44}$$

Following the SIMO FIR channel results of [16], (44) can be used to to estimate the MIMO channel impulse response $\mathbf{F}_i$ for $i = 0, 1, \cdots, D-1$ (for arbitrary D) up to a unitary matrix ([1]). [This unitary matrix requires higher-order statistics for its estimation [1].] Knowledge of $\mathbf{F}_i$ for $i = 0, 1, \cdots, D-1$ can be used to design MMSE equalizer with lag $D-1$ [10]. All of this relies on **(H4)** which allows determination of the noise variance from data by eigendecomposition of the data correlation matrix (which is discussed next).

4. ESTIMATION OF NOISE VARIANCE

In practice, we have noisy measurements $\mathbf{y}(k)$ of $\mathbf{s}(k)$ whereas the preceding discussion and results are based upon availability of the correlation function of $\mathbf{s}(k)$. Lemma 1 below is useful in this regard. Consider the case of $l = 1$ (one-step prediction). By (42) we have

$$\mathbf{s}(k) = -\sum_{i=1}^{L_1} \overline{\mathbf{A}}_{1i}\mathbf{s}(k-i) + \mathbf{F}_0\mathbf{w}(k). \tag{45}$$

If $L_1 > n_a + Mn_b$, then $\overline{\mathbf{A}}_i = 0$ for $i > n_a + Mn_b$ by virtue of (25).

$\boxed{\textbf{Lemma 1.}}$ Under **(H1)-(H3)** and **(H5)**, $\text{rank}\{\mathcal{R}_{ssL_1}\} \leq NL_1 + M$ for $L_1 \geq n_a + Mn_b$ where $\mathcal{R}_{ssL_1}$ is $[N(L_1+1)] \times [N(L_1+1)]$ with its ij−th block-element given by $\mathbf{R}_{ss}(j-i) := E\{\mathbf{s}(k+j-i)\mathbf{s}^H(k)\}$. •

Sketch of proof: It follows from (45) and the fact $\mathbf{F}_0 = \mathbf{B}_0$ that

$$\begin{bmatrix} I_N & \overline{\mathbf{A}}_1 & \cdots & \overline{\mathbf{A}}_{L_1} \end{bmatrix} \mathcal{R}_{ssL_1} = \begin{bmatrix} \mathbf{B}_0\mathbf{B}_0^H & 0 & \cdots & 0 \end{bmatrix} \tag{46}$$

Clearly $\text{rank}\{\begin{bmatrix} I_N & \overline{\mathbf{A}}_1 & \cdots & \overline{\mathbf{A}}_{L_1} \end{bmatrix}\} = N$. By **(H2)**, $\text{rank}\{\mathbf{B}_0\} = M = \text{rank}\{\mathbf{B}_0\mathbf{B}_0^H\}$. The desired result then follows from (46) and the Sylvester's inequality [5, p. 655]. □

In a fashion similar to $\mathcal{R}_{ssL_1}$ in Lemma 1, let $\mathcal{R}_{yyL_1}$ denote a $[N(L_1+1)] \times [N(L_1+1)]$ matrix with its ij-th block element as $\mathbf{R}_{yy}(j-i) = E\{\mathbf{y}(k+j-i)\mathbf{y}^H(k)\}$; define similarly $\mathcal{R}_{nnL_1}$ pertaining to the additive noise. Carry out an eigendecomposition of $\mathcal{R}_{yyL_1}$. Then the smallest $N-M$ eigenvalues of $\mathcal{R}_{yyL_1}$ equal σ_n^2 because under **(H1)-(H3)** and **(H5)**, $\text{rank}\{(\mathcal{R}_{ssL_1}\} \leq NL_1 + M$ whereas under **(H4)**, $\text{rank}\{\mathcal{R}_{nnL_1}\} = NL_1 + N = \text{rank}\{\mathcal{R}_{yyL_1}\}$. Thus a consistent estimate $\widehat{\sigma}_n^2$ of σ_n^2 is obtained by taking it as the average of the smallest $N-M$ eigenvalues of $\widehat{\mathcal{R}}_{yyL_1}$, the data-based consistent estimate of $\mathcal{R}_{yyL_1}$. The noise-free signal correlation function can then be estimated from the noisy-data correlations.

5. REFERENCES

[1] A. Gorokhov, P. Loubaton and E. Moulines, "Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach," in *Proc. 1996 ICASSP*, pp. 2415-2418, Atlanta, GA, May 7-10, 1996.

[2] K. Abed-Meraim, E. Moulines and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 694-705, March 1997.

[3] Special Issue, *IEEE Trans. Signal Processing*, vol. SP-45, Jan. 1997.

[4] P.E. Caines, *Linear Stochastic Systems.* New York: Wiley, 1988.

[5] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[6] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures," in *Proc. 1996 ICASSP*, pp. 2940-2943, Atlanta, GA, May 7-10, 1996.

[7] A. Gorokhov and P. Loubaton, "Multiple input multiple output ARMA systems: second order blind identification for signal extraction," in *Proc. IEEE SSAP Workshop*, pp. 348-351, Corfu, Greece, June 1996.

[8] J.K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Trans. Signal Processing*, vol. SP-45, pp. 268-271, Jan. 1997.

[9] R. Bitmead et al., "Greatest common division via generalized Sylvester and Bezout matrices," *IEEE Trans. Autom. Control*, vol. 23, pp. 1043-1047, Dec. 1978.

[10] J.K. Tugnait, "FIR inverses to MIMO rational transfer functions with application to blind equalization," in *Proc. 30th Annual Asilomar Conf. Signals Sys. Comp.*, pp. 295-299, Pacific Grove, CA, Nov. 1996.

[11] B. Huang and J.K. Tugnait, "Blind equalization of I.I.R. single-input multiple-output channels with common zeros using second-order statistics," in *Proc. 1998 IEEE ICASSP*, Seattle, WA, May 12-15, 1998.

[12] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction," in *Proc. 1994 IEEE ICASSP*, pp. IV:585-588, Adelaide, Australia, May 1994.

[13] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channels," *IEEE Trans. Information Theory*, vol. IT-40, pp. 340-349, March 1994.

[14] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Proc.*, vol. SP-43, pp. 516-525, Feb. 1995.

[15] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Proc.*, vol. SP-45, pp. 3053-3061, Dec. 1997.

[16] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. 1997 ICASSP*, pp. 3621-3624, April 21-24, 1997.